Vol. 56, No. 4 July, 1949

Psychological Review

EDITED BY
CARROLL C. PRATT
PRINCETON UNIVERSITY

CONTENTS

An Interpretation of Pseudo-Conditioning: MARK A. MAY 117

Emotion as Disorganised Response—A Reply to Professor Leeper: PAUL THOMAS YOUNG 184

The Psychology of Invention in a Very Simple Case: EDWARD L. THORNDIKE 192

Value, Need, and Other Factors in Perception: LAUNOR F. CARTER AND KERMIT SCHOOLER 200

A Theory Relating Momentary Effective Reaction Potential to Response Latency: C. J. BURKE 208

Color Vision and Factor Analysis: JOZEF COHEN 224

Homeostasis and Servo-Mechanisms for What? CHRISTIAN O. WEBER .. 234

PUBLISHED BI-MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

$5.50 volume $1.00 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. and R., authorized Jan. 8, 1948

COMPLETE MEMBERSHIP LIST

Approximately 8,000 names

The American Psychological Association maintains an address list of its members and affiliates, which is for sale providing the nature of its use is in conformity with the purposes of the Association.

Envelopes addressed $30.00
(advertiser furnishes envelopes and pays express charges)

Addresses on tape, not gummed $15.00
(suitable for a mailing machine)

STATE LISTS

Priced according to number of names wanted

SUPPLEMENTARY LISTS

Approximately 3,500 names in total list
Individual journal lists vary from 400 to 1,600

The Association also maintains a list of subscribers who are not members of the Association (universities, libraries, industrial laboratories, hospitals, other types of institutions, and individual subscribers). The general list for all journals includes all types. Each single journal has a more specialized circulation.

For any one journal, envelopes addressed $15.00
For any one journal, addresses on tape $10.00
For all journals, envelopes addressed $20.00
For all journals, addresses on tape $15.00

For further information, write to

American Psychological Association
1515 Massachusetts Avenue Northwest
Washington 5, D. C.

VOL. 56, NO. 4 JULY, 1949

THE PSYCHOLOGICAL REVIEW

AN INTERPRETATION OF PSEUDO-CONDITIONING

BY MARK A. MAY

Institute of Human Relations,
Yale University

The phenomenon known as pseudo-conditioning is of considerable theoretical interest because it appears to contradict the important principle of temporal contiguity between stimulus and response in conditioning. The term was coined, I believe, by Grether (6), whose experiments with monkeys well indicate how it is obtained. In conditioning fear responses to an electric bell he used, as unconditioned stimuli, flashlight powder and the uncurling of an artificial snake. In a control group these unconditioned stimuli were presented ten times alone, *i.e.*, without pairing either with the bell. On the eleventh and subsequent trials the bell alone evoked fear responses quite similar to those produced by the unconditioned stimulus. He found further that any sudden loud noise would evoke similar responses.

Several experimenters have investigated this phenomenon. It has been observed in adult humans by Grant and Hilgard (4), Grant and Meyer (5), and Harris (10); in infants by Wickens and Wickens (17); in cats by Harlow and Toltzien (8); in rats by Harris (9; 11), May (13), and Wickens and Wickens (18); and in goldfish by Harlow (7) and Sears (15). However, the phenomenon is not universal in the sense that all subjects in all experiments show it, though it can be obtained fairly easily under proper experimental conditions.

The experimental procedure usually followed is: (1) the subject is pretested with a neutral stimulus to discover whether or not it will evoke any response resembling the reactions evoked by the unconditioned stimulus; if it does, the subject is either discarded or habituated to the point of neutralization of the stimulus. (2) The unconditioned stimulus is applied for a number of trials (ten or more) and the responses to it are recorded. Sometimes escape training is given so that one response dominates all others. (3) The subject is now again tested with the neutral stimulus; if the response is the same as or similar to those to the unconditioned stimulus, the trial is recorded as an instance of pseudo-conditioning. Investigators are not wholly agreed on what kinds of responses should be counted. Some include all that are qualitatively similar to those of the unconditioned stimulus; others have used quantitative scales (8).

The experimental work that has been reported on pseudo-conditioning is rather exploratory but suggests certain functional relations between experimental conditions and the probabilities of evoking pseudo-conditioned

responses.[1] Although none of these relations has been systematically determined, they are stated below as tentative empirical generalizations. The experimental evidence in support of each is summarized.

1. *The probability that a previously neutral stimulus will evoke a PCR is a function of the degree of similarity between it and the unconditioned stimulus that was used during the pseudo-conditioning trials.* By *similarity* is meant either the same sense modality or some formal relation such as suddenness of onset.

An example of primary stimulus-generalization on an intensity gradient of the same sense modality was found by Harlow (7). Eight goldfish were pretested for fear responses with a very weak (liminal) shock and with an audio-vibrator. Each was then given ten strong shocks a day at intervals of 100 to 130 seconds, to which they responded with "violent struggling and flight responses." After two days of training they were retested with liminal shock and with audio-vibration. PCR's were obtained to both stimuli. The liminal shock is a case of primary stimulus-generalization on an intensity gradient. The audio-vibration may be similar to shock by some partial identity or quality such as oscillation. One does not know, of course, how a strong electric shock feels to a goldfish but we do know that in humans it produces a tingling sensation similar to that caused by a tactile vibrator such as is sometimes used by barbers.

A typical Pavlovian gradient of generalization was found by Grant and Dittmer (3) in human subjects. The unconditioned stimulus was shock, and the unconditioned response was the galvanic skin reaction. The pseudo-conditioned stimulus was a vibro-tactile sensation produced by a Hull-Bass stimulator. Preliminary tests with the stimulator were given at four points on the subject's back at varying distances from the point at which shock was later applied. Following these pretests, each subject was given five shocks at 10–30 second intervals on one point on the back. These were followed by a series of tests with the tactile stimulus applied at points of varying distance from the point of shock. Means and standard deviations of GSR's at point of shock and at three points of varying distance from it were:

Point of Shock	1	2	3
3.33 ± .55	3.12 ± .65	3.04 ± .69	2.34 ± .48

The authors concluded that the "average galvanic skin responses of 15 subjects to vibro-tactile stimulators placed along the back gave a clear-cut gradient of response amplitude after the presentation of a number of electric shocks."

Wickens and Wickens (18) have reported an experiment demonstrating an instance of *formal similarity* between the neutral and the unconditioned stimulus. The unconditioned stimulus was shock in a shuttle-box situation in which rats ran alternately from one end of the box to the other to escape the shock; the conditioned stimulus was a two-volt pilot light bulb. The apparatus was arranged in such a way that either the conditioned or the unconditioned stimulus could be applied *suddenly* or *gradually*. Thirty-seven animals were trained and tested in four different groups. In the first group the onset of shock was *sudden* on the training trials and the onset of light *sudden* on the test trials. In the second group, shock was *gradual* on the training trials and light was *gradual* on the test trials.

[1] Hereafter abbreviated as PCR.

Seven out of nine animals in the first group, and eight out of ten in the second group gave PCR's. In groups 3 and 4, the onset of shock was the opposite from the onset of the light. In these two groups only one out of nine and two out of nine, respectively, gave PCR's. The authors state in summary: "These results are consistent with the hypothesis that pseudo-conditioning may be akin to Pavlovian stimulus generalization, and is not in accord with the hypothesis that pseudo-conditioning arises from an attitude of expectancy." This conclusion would have been more convincing had the animals been tested with a wider range of stimuli.

2. *The probability that a previously neutral stimulus will evoke a PCR may be a function of the similarity between the environment in which the pseudo-conditioning trials are given and that in which the test trials occur.*

The experimental evidence on this point is meager. Harlow and Toltzien report that in their experiment with cats (8, p. 370), "PCR's appear more frequently in the experimental situation than in the neutral environment. PCR's appeared 75 times in the experimental situation, 54 times in the neutral environment. In only one subject, No. 8, did PCR's appear in the experimental situation and fail to appear in the neutral environment. The average intensity rating of the responses in the experimental situation was 5.5, in the neutral environment, 5.0. The difference probably would have been greater had not the test trials always been presented first in the neutral environment."

3. *The probability that a previously neutral stimulus will evoke a PCR appears to be a function of the number of pseudo-conditioning trials given previous to the test trials.*

The evidence for this is again from Harlow and Toltzien, who report as follows (8, p. 370): "After 10 presentations of the US, PCR's appeared 14 times. After 20 presentations of the US, PCR's appeared 57 times. Ten additional presentations had little effect on the frequency level; PCR's then appeared 58 times. A maximum of 130 PCR's was possible in each case."

4. *The probability that a previously neutral stimulus will evoke a PCR may be a function of the length of the time elapsing between the last pseudo-conditioning trials and the test trials.*

The evidence on this point is conflicting. Harlow and Toltzien tested some of the cats after five minutes, some after three hours, and some after 24 hours, and found no significant differences in the percentages of the PCR's. The writer has some fragmentary data on rats which indicate that there is a decreasing tendency for PCR's to occur as the time between shock trials and test trials is increased.

5. *The probability that a previously neutral stimulus will evoke a PCR may be a function of the strength of the unconditioned stimulus which is used on the pseudo-conditioning trials.*

In most of the experiments reported on pseudo-conditioning, medium to strong electric shock was the unconditioned stimulus but, to the knowledge of the writer, no one has systematically varied the strength of shock (keeping other variables constant) and noted resultant variations in occurrence, strength, and latency of PCR's.

There is direct evidence, however, on the relationship between the strength of the unconditioned stimulus and the acquisition of further conditioned responses. Passey (14) has reported a study in which four intensities of air pressure were used in producing an air-puff stimulus to the eyeball in an eyelid conditioning experiment. He found that the rate of acquisition of the conditioned response is proportional to the

logarithm of the air pressure of the puff.

6. *The probability that a previously neutral stimulus will evoke a PCR may be a function of the intensity of the test stimulus.*

There is no direct evidence on this point. It is mentioned because in certain types of experiments, as, for example, the shuttle-box, the environmental cues may acquire conditioned inhibition. In this case the test stimulus may serve as a disinhibitor of the conditioned inhibition. It has been observed by Pavlov, and later by Hull and others, that a relatively weak stimulus is a more effective disinhibitor than a stronger one. Thus one might suppose that if the test stimulus in pseudo-conditioning is acting as a disinhibitor, it would be more apt to evoke a PCR if it is relatively weak.

This completes our brief survey of the experimental conditions that may affect the occurrence, strength, latency, and duration of PCR's. Before proceeding to a theoretical interpretation of pseudo-conditioning, a glance at some of the facts concerning the relationship between pseudo-conditioning trials and subsequent forward conditioning may be in order.

There is ample evidence in the existing experimental literature that preliminary applications of the unconditioned stimulus alone have definite effects on the rate and amplitude of subsequent forward conditioning. Long ago Beritoff (1) found that if a dog is given a few shocks before the conditioning trials begin, the rate of conditioning is accelerated. Girden (2) reported an experiment with dogs in which it appears that the number of applications of the unconditioned stimulus alone, prior to forward conditioning, has a qualitative effect on the character of the subsequent conditioned responses. Harris (11) has reported some interesting data on the effects of UCS trials on later buzzer-shock conditioning. Using a stabilimeter, he measured the amplitude of gross responses to shock alone and later to buzzer-shock combination. In one experiment he systematically varied the number of UCS trials and measured their effect on amplitude of later buzzer-shock trials. The first group of rats received one day of conditioning training only, being given no UCS trials at all; the second group had 10 massed UCS trials in one day; the third had 10 a day for two days, and so on until finally the seventh group had 10 a day for 10 days. All groups received 10 massed buzzer-shock trials on the day following the last shock-alone trials. When the mean amplitude of response made by the first group in the buzzer-shock trials was compared with the mean amplitudes of response made by the various UCS groups, the following results were obtained:

Days of UCS trials	1	2	3	5	7	10
Mean amplitudes of response of UCS groups	31.0	24.7	25.7	25.7	26.3	32.4
Mean amplitude of response of conditioning group	17.6	17.6	17.6	17.6	17.6	17.6
Difference	13.4	7.1	8.1	8.1	8.7	14.8

Harris suggests that the drop in amplitude from the first to the second day may be attributed to apparatus habituation.

Most of the papers on pseudo-conditioning conclude with a paragraph

or two on interpretation. The explanatory hypotheses that have been advanced may be classified as "associative" and "non-associative." The associative hypotheses attempt to explain the phenomenon on the grounds of primary stimulus-generalization, conditioned trace stimuli, or situational conditioning, and claim that it does not violate the principle of temporal contiguity. The non-associative hypotheses suggest that it may be due to *sensitization, response dominance,* a *generalized motor set,* or *expectancy.*

The *sensitization* hypothesis asserts that the frequent application of a strong stimulus, such as shock, has a cumulative effect on the nervous system of the animal and tends to lower all response thresholds, thereby making the subject more responsive to all stimuli. The *dominance* hypothesis asserts that a response becomes dominant when it is repeated frequently for several trials, when each trial is reinforced by escape from a noxious stimulus, and when no competing responses are reinforced meanwhile. During the time a response is dominant, it tends to supersede all other responses which may be made in that environment. The *expectancy* hypothesis assumes that the repeated experience of being shocked generates in the subject a generalized motor set of readiness for shock. Thus if the animal is "expecting" to be shocked, any noticeable change in the situation will evoke some kind of an avoidance response. While none of these non-associative interpretations seems adequate to account for the facts, each represents concepts that are useful in formulating a more comprehensive theory. A theory is needed in which both associative and non-associative factors are combined.

Hull's theory of behavior, with one modification, will, I think, account for all the facts of pseudo-conditioning. It includes associative and non-associative factors which correspond closely to the explanatory hypotheses mentioned above, with the exception that it contains no factor representing response dominance, or facilitation.

Hull's basic formula for the effective excitatory reaction potential of a stimulus for a response is (12):

$$_{S}\bar{E}_{R} = (_{S}\bar{H}_{R} \times D) - (_{S}I_{R} + I_{R}),$$

where $_{S}\bar{H}_{R}$ is the effective habit strength of the stimulus for the response (*i.e.,* the innate strength plus that gained from primary stimulus-generalization); D is the drive; $_{S}I_{R}$, conditioned inhibition; and I_{R}, the reactive inhibition. Applied to pseudo-conditioning, drive is the sensitization produced by the after-effects of shock and/or the acquired fear of the situation. These effects presumably persist for awhile, fading out according to some unknown law of decay. $_{S}\bar{H}_{R}$ is the positive habit strength between the test stimulus and the PCR. It includes both the innate strength which is presumably very weak, and the accumulated additional strength gained by primary stimulus-generalization from the unconditioned stimulus. In addition to the neutral or test stimulus there are, of course, the environmental cues which, as indicated above, may have acquired some positive habit strengths with the PCR and also a certain amount of conditioned inhibition. The conditioned inhibition is presumably the stronger of the two in test situations where the animal does not give a PCR to the environmental cues alone.

On the test trials, when the neutral stimulus is applied, two things happen. First, the stimulus disinhibits the conditioned inhibition of the environmental cues and somewhat weakens the total value of the inhibitory potential. Second, the drive (the sensitization effect or the acquired drive of fear) in-

creases the weak habit strength of the neutral stimulus to the point where the total excitatory potential is greater than the inhibitory. This state of affairs should account for the evocation of the PCR, provided the excitatory potential is sufficiently above the threshold as not to be affected by the oscillation factor. If the response occurs and is reinforced by drive reduction, ${}_{S}\bar{H}_{R}$ will gain in strength and finally reach the point where S will evoke PCR even though the drive is weak.

So far this interpretation contains a weakness. According to Hull's theory, drive acts as an indiscriminate multiplier of all habit strengths between a given stimulus and all of the responses with which it is connected (12). In itself primary drive is not selective but merely increases the general responsiveness of the organism. The neutral stimuli used in pseudo-conditioning experiments may have, and probably do have, connections with other responses of greater habit strengths than those of the PCR's.

Here is where response dominance or set comes in. A response which has just been evoked a number of times, so that the animal is warmed up, appears in many instances at least to gain an advantage over competing responses. This warming-up effect has been noticed in a variety of experiments (see, for example, 16, Fig. 10), and is particularly apparent in massed training where several trials are given in one session. The first few trials at the beginning of each day's training are often less successful than the succeeding ones of that day. In two different experiments by the author it was observed that *one* warming-up trial is quite sufficient to overcome what appears to be "initial inertia" of the animal.

The mechanism underlying response dominance and the warming-up effect could depend on any one of several possibilities. One is that the stimulus traces of the initial response are conditioned to subsequent responses of the same kind. These traces combined with the environmental cues and the "test" stimulus have sufficient habit strength to evoke the response. But whatever the mechanism may be, the phenomenon itself could be called "reactive facilitation." In Hull's formula it clearly belongs with and forms a part of the *excitatory potential.* Just how it may be combined with the factors of drive (D) and *effective habit strength* (${}_{S}\bar{H}_{R}$) is a question that cannot be answered at the moment. Suppose, however, that it is simply added. Hull's formula then might be like this:

$$ {}_{S}\bar{E}_{R} = D({}_{S}\bar{H}_{R} + F_{R}) - ({}_{S}I_{R} + I_{R}), $$

in which F_R is the factor of reactive facilitation. Stated in some such fashion as this, the formula includes all of the associative and non-associative factors necessary to account for the experimental results thus far obtained.

The test of the validity of this interpretation is, of course, whether or not deductions from it can be experimentally verified. It would predict that the probability of getting a PCR on the first test trial is a function of (a) the drive strength of the unconditioned stimulus, (b) the time interval between the last application of the unconditioned stimulus and the first test trial, (c) the number of applications of the unconditioned stimulus (up to a certain point), (d) the distribution or spacing of the unconditioned trials, and (e) the degree of similarity between the unconditioned stimulus and the test stimulus.

SUMMARY

The phenomenon of pseudo-conditioning appears to contradict the important principle of temporal contiguity between stimulus and response in con-

ditioning, but it is well established experimentally. A neutral stimulus that will not evoke a response on a pretest, will evoke it later provided it has just been evoked a number of times by a fairly strong unconditioned stimulus with which the test stimulus *has not been paired*. This phenomenon has been attributed by various experimenters to such non-associative factors as sensitization, response dominance, generalized motor set and expectancy; and by others to associative factors such as stimulus generalization and disinhibition. The writer has attempted to show how the experimental results may be accounted for theoretically by adding a factor of "reactive facilitation" to Hull's basic formula for *effective reaction potential*.

REFERENCES

1. BERITOFF, J. Ueber die individuell-erworbene Tätigkeit des Zentralnervensystems bei Tauben. *Arch. f. d. ges. Physiol.*, 1926, **213**, 370–406.
2. GIRDEN, E. Conditioning and problem solving behavior. *Amer. J. Psychol.*, 1938, **51**, 677–687.
3. GRANT, D. A., & DITTMER, D. G. A tactile generalization gradient for a pseudo-conditioned response. *J. exp. Psychol.*, 1940, **26**, 404–412.
4. ——, & HILGARD, E. R. Sensitization as a supplement to association in eyelid conditioning. *Psychol. Bull.*, 1940, **37**, 478–479.
5. ——, & MEYER, H. I. The formation of generalized response sets during repeated electric shock stimulation. *J. gen. Psychol.*, 1941, **24**, 21–38.
6. GRETHER, W. F. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *J. comp. Psychol.*, 1938, **25**, 91–196.
7. HARLOW, H. F. Forward conditioning, backward conditioning, and pseudo-conditioning in the goldfish. *J. genet. Psychol.*, 1939, **55**, 49–58.
8. ——, & TOLTZIEN, F. Formation of pseudo-conditioned responses in the cat. *J. gen. Psychol.*, 1940, **23**, 367–375.
9. HARRIS, J. D. An analysis of certain non-associative factors inherent in avoidance conditioning in the rat. *Psychol. Bull.*, 1941, **38**, 572.
10. ——. Forward conditioning, backward conditioning, pseudo-conditioning, and adaptation to the conditioned stimulus. *J. exp. Psychol.*, 1941, **28**, 491–502.
11. ——. Studies in non-association factors inherent in conditioning. *Comp. Psychol. Monogr.*, 1943, **18**, No. 1.
12. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century Co., Inc., 1943.
13. MAY, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.
14. PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
15. SEARS, R. R. Effect of optic lobe ablation on the visuomotor behavior of goldfish. *J. comp. Psychol.*, 1934, **17**, 233–265.
16. WHITING, J. W. M., & MOWRER, O. H. Habit progression and regression—a laboratory study of some factors relevant to human socialization. *J. comp. Psychol.*, 1943, **36**, 229–253.
17. WICKENS, D. D., & WICKENS, C. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, **26**, 94–102.
18. ————. Some factors related to pseudo-conditioning. *J. exp. Psychol.*, 1942, **31**, 518–526.

[MS. received November 15, 1948]

EMOTION AS DISORGANIZED RESPONSE—A REPLY TO PROFESSOR LEEPER [1]

BY PAUL THOMAS YOUNG

University of Illinois

The view that emotion is a disorganized (or disorganizing) response is held quite commonly among psychologists today. This interpretation of emotion is found in the textbook writings of Munn, Shaffer *et al.*, Young, Landis (in the text of Boring, Langfeld, and Weld), Dockeray, Woodworth, and others. This view, Professor Leeper (1) writes, is such a poor generalization that it has served as a serious obstacle to good research and to application.

The disorganization doctrine, he tells us, rests upon undefined and wrongly interpreted terms. None of the above writers has attempted to give a definition of the terms *disorganized* and *disorganizing*. These key terms must be defined to prevent sloppy thinking and the neglect of significant data. In fact, some of the above writers have used the concept of disorganization in shifting and contradictory meanings.

To help remedy this deplorable condition Professor Leeper gives a definition of the key terms *organize* and *disorganize*. A system is organized, he writes, when one part of it is functioning harmoniously with the other parts. A system is organized when the parts fit, or dovetail, or are congruous with one another. Something is disorganized when the parts do not function harmoniously but work at cross purposes with each other.

Now in anger, to illustrate, the person is definitely organized for attack. His visceral processes are organized for the prolonged and vigorous action required in a biological emergency. His conscious experience is organized along with the basic determination. Instinctive acts of attack or flight or mating are highly organized—viscerally, behaviorally, and in conscious experience.

Again, according to Leeper, a neurotic individual is not completely disorganized. He may be organized around a feeling of worthlessness which is so rigid that he cannot shake it off or around a feeling of discouragement or anger or what not. Organization rather than disorganization is the rule in "emotion."

Professor Leeper admits, however, that disorganization does occur in extreme emotions. A man may be so extremely afraid that he wants to run or to call for help but he may be unable to do so. The boxer may become so angry that he loses his technical skill and fights ineffectively. In these extreme cases "the emotional process does seem to produce disorganization in significant degree. It interferes with reactions that would be consistent or harmonious with what is demanded by the emotion." But, says Professor Leeper, these are extreme cases and we should not use them as a means of deciding the normal function of emotional responses in the various degrees of intensity. A concept should not be defined by the extremes. The concept of emotion must apply to all degrees and intensities of emotion.

Leeper is clear about this point: Emotion exists in all degrees of intensity, but disorganization character-

[1] In the preparation of this paper the writer has had helpful criticisms from S. S. Dubin, Stanley Roscoe, and R. G. Smith, of the University of Illinois.

izes only the most extreme or intense emotions. Moderate emotions are *organized* responses. Hardly in any sense, he writes, do we find disorganization in affection, in esthetic emotion, in friendliness, in sympathy. These "emotions," as well as anger and fear, *organize* the individual for action.

"Emotions" operate primarily as motives, Professor Leeper continues, and their analysis belongs within the context of motivation. In contrast with the disorganization theory of emotion, Leeper affirms, emotional processes of all sorts (except perhaps in the rare and intense forms) are organizing in their influence and they should be studied as an aspect of the motivation of the higher animals. In fear or anger, to repeat, the person is organized so that he can most naturally do something consistent with his fear or anger.

Emotional motivation rests upon relatively complex and primitive neural patterns. From the evolutionary point of view animals motivated solely by physiological drives were not well equipped for survival. The animal that did not make avoiding responses until he was grabbed by an enemy was less likely to survive than an animal capable of reactions of fear set off by relatively slight stimuli.

Emotional motivation is indispensable in human life. Emotion is not merely a disturbance or a disorganization.

CRITIQUE OF EMOTIONAL ORGANIZATION AND DISORGANIZATION

Professor Leeper's discussion is interesting and important. He is struggling with a genuine difficulty in the definition of emotion—a difficulty which pervades the whole of affective and dynamic psychology. Terms must be defined, as Leeper has pointed out, so let us press this matter of definition a little further in the light of Leeper's paper. We believe that the difficulty is entirely one of definition.

The first point to notice is that Leeper has not given us a precise definition of the key terms *emotion* and *motivation,* although his paper bears the title "A motivational theory of emotion . . ." and stresses the importance of defining key terms.

Leeper emphasizes the point that an "emotion" is typically an *organized* and *organizing* process, but he admits that in extreme or intense "emotions" there may be disorganization. He states that the definition of a key term in psychology must cover the phenomenon as it appears in all degrees and not merely at the extremes; yet he does not give us a definition of emotion which includes both organized and disorganized "emotions."

Leeper's admission that intense emotions are disorganized psychological processes raises a curious question of logic. Logically the concepts of organization and disorganization are opposed and both concepts are not necessary to interpret the facts. The greatest possible disorganization is at the same time the minimum of organization and *vice versa.* If we hold to the concept of organization exclusively, then these extreme or intense emotions are processes with the minimum of organization. The degree of organization is so low, however, that we find ourselves spontaneously using such words as *upset, disturbance, disorganization, disruption,* to describe them adequately.

The attempt to classify emotions into two groups, organized and disorganized, is futile since no sharp line of distinction can be drawn. The true distinction, rather, is between two aspects of emotional behavior. In his book upon emotion (2, p. 176) the present writer has described the relationship in these words:

"A true picture of the relation between organized and disorganized behavior is this. At a given time the behavior of an individual is more or less organized, and more or less disorganized. One can view the same bit of behavior from the point of view of its integration or from the obverse standpoint of disintegration. The first point of view reveals an organized, integrated aspect; the second, a disorganized, disintegrated aspect.

We shall use the phrase *emotional behavior* to designate the total process regardless of the standpoint from which it is viewed. The adjective *emotional* implies that the behavior is to some extent disorganized, disintegrated, upset, disturbed. This is the aspect in which we are interested when we are studying *emotion*. But, when examined from another point of view, emotional behavior is found to contain well-organized components—integrated patterns of response, goal-oriented activities. . . ."

Organization and disorganization are thus different ways of viewing and interpreting one and the same event. Even the most completely disorganized emotional process, when viewed from another point of view, is found to contain integrated components of response. There is no real antagonism between the two views of emotional behavior. As stated above, they are simply different ways of interpreting the same facts. Both interpretations are needed and both are legitimate.

Leeper's view that only abnormal or extreme emotions are disorganized and that normal emotions are organized is not an adequate statement of the facts. Starting with the logical opposition of these two words, we are unable to reach the conclusion that psychological processes are of two kinds—organized and disorganized. As a matter of fact only one of these concepts is required logically and any distinction between organized and disorganized is relative and arbitrary.

The confusion lies in the logical opposition of words. Emotional behavior, as an event in nature, simply occurs. It is the psychologist who, with one interest, emphasizes now the aspect of organization and again, with another interest, the aspect of disorganization. But to force the psychologist to choose between these interpretations is ill advised and confusing.

The writer agrees with Leeper in emphasizing the importance of the organized components of emotional behavior. Organized patterns of emotional response include: (a) reflexive patterns such as the rage pattern, the cry, the laugh, the mating patterns, etc., (b) persistent purposive activities such as flight, attack, courtship, etc., (c) culturally determined gestures and postures such as the attitudes of courage (fear-inhibitory), friendliness (inhibitory of hostile attack), the voluntary smile, etc.

Organized patterns assuredly exist as components of emotional behavior. They are important practically and their analysis, as part of the psychology of motivation, must not be neglected. The present writer, certainly, has given adequate emphasis to the organized processes occurring in emotional behavior (2). Further, in a current paper, based upon experimental findings, he has argued that the enjoyment of food and the relief from distress *organize* determinations to seek specific kinds of food (3). This frankly hedonic theory of motivation emphasizes the positive role of enjoyment and relief from distress in organizing purposive behavior.

There is no incompatibility between Professor Leeper's emphasis upon organized response in emotional behavior and our own account. Nor is there anything really novel about Leeper's emphasis since other students of emotion, *e.g.*, McDougall, have stressed integrated behavior.

The present writer has argued that the analysis of organized patterns of behavior, and especially goal-oriented behavior, belongs squarely within the psychology of motivation. This is just as true in explaining a race for one's life as in interpreting persistent food-seeking behavior. But the psychologist who would limit his view of emotional behavior to the organized, integrated, aspect alone is thereby closing his eyes to another important aspect: the obverse aspect of disorganization.

Whenever a highly organized activity is blocked, or when success with its release of tension is suddenly achieved, or when some danger threatens, there is an immediate affective upset. Emotional disorganization is a condition which exists in nature as truly as smooth and calm organized activity. There is no gain in ignoring this aspect of the total picture.

EMOTION AND MOTIVATION

In an all too brief positive statement toward the close of his discussion Professor Leeper states that emotions *are* motives and that the analysis of emotional behavior should be made within motivational psychology.

From our point of view an acute affective upset is not itself a motive. An affective upset, however, may result in the formation of a motive which has consequences in future behavior. To illustrate: According to the record, when Abraham Lincoln saw slaves being sold on the New Orleans market he was so disturbed emotionally that he resolved "to hit that thing hard" if ever the opportunity came. This resolution, resulting from an emotional disturbance, was a determining factor in Lincoln's life and in human destiny. The upset, a disorganized event, did not motivate but it led to the formation of a motive.

Now insofar as emotional behavior can be regarded as an *organized* process its analysis, we fully agree, belongs within motivational psychology. In emotional behavior well integrated patterns of response appear. Why not study the motivation of these organized responses exactly as one studies reflex action, behavioral drive, social expression, and other smoothly organized activities? The writer's answer is: Go ahead! Get all the light possible upon these and other integrated activities. Why not?

We should prefer to say, however, that these integrated patterns of response are *motivated* rather than that they are *motives*. Motives are factors which initiate, direct, and sustain action. Disorganized affective states are produced by the frustration of motives, by conflict, by intense and persistent painful stimulation, by the sudden release of tension as in joyful success, by anticipation of harm, and in other ways. These acute affective disturbances reflect the dynamic interplay of motives.

After the story of motivation has been told completely there will remain facts of a different order to be investigated. Acute affective disorganization exists within the individual as a fact of nature. We cannot, therefore, agree with Professor Leeper that the concept of disorganization serves as a serious obstacle to good research and to application. Good research and application must rest first of all upon a true view of the facts.

EMOTION AND OTHER AFFECTIVE PROCESSES

Another point relative to the definition of emotion is the following:

In popular psychology the word *emotion* is used broadly to cover a wide range of affective and motivational processes such as persistent states of anxiety, phobias, attitudes of love and

hate, determinations to seek revenge, and the like. This popular usage ignores two important distinctions: (1) the distinction between affective processes in general and a specific form of affective process which is characterized by upset, and (2) the distinction between emotion as an event and an assumed stable state of the subject. These two distinctions will be considered briefly below:

(1) The broad concept of affective process includes at least: (a) simple feelings of pleasantness and unpleasantness, (b) moods of cheerfulness and depression, (c) interests and aversions, (d) organic aches and satisfactions, (e) feelings of hunger, thirst, fatigue, etc., and (f) emotional upsets. The present writer has distinguished six main varieties of affective process, determined by their conditions, and he identified emotion with only one, namely, that form of affective process characterized by upset, or acute affective disturbance (2, pp. 271–273). Of course, if emotion is defined as an affective process characterized by acute disturbance, the term must mean exactly that and it must be used consistently in that rather limited and specific sense.

The popular term *emotion,* however, has not been so sharply delimited. It includes moods, sentiments, interests, and other affective processes, in which the factor of upset is definitely less pronounced than with the truly disruptive emotions. The popular term *emotion* clearly covers some affective processes in which organized behavior is very prominent. This broad use of the term *emotion* may be a source of confusion.

To clarify the picture it is fair to ask Professor Leeper these questions: Is *emotion* to be regarded as the equivalent of affective process, in the broad meaning of affectivity? If not, how are emotions to be distinguished from other kinds of affective process?

Our answer would be that emotion is a specific kind of affective process, namely, that kind of affective process which is characterized by acute upset or disturbance involving particularly the smooth muscles and glands. We suggest that the phrase *affective process* be used broadly to include simple feelings, moods, sentiments, interesting and aversive activities, emotions, and similar processes; and that the term *emotion* be restricted to those affective processes characterized by acute upset or disorganization. This suggestion implies that the popular meaning of emotion should be narrowed down to the meaning of acute affective upset or disturbance. Where we now speak of emotional development, emotional maturity, emotional appeal, and the like, we should substitute the more technically correct term *affective* for *emotional.*

(2) There is a further difficulty of definition. As an *observed* event emotion must be distinguished from the *assumed* event within the subject and from his relatively stable affective states. As a psychological datum emotion is revealed in direct conscious experience (conscious emotion), in behavior (emotional behavior), and in internal bodily changes especially in those bodily processes regulated through the autonomic nervous system (physiological emotion). Now an emotion is *assumed* to be an acute affective disturbance within the individual as a whole.

The word *acute* implies that the emotion is brief and intense. In common with other acute events it has a beginning, a relatively brief course, and an ending. The acute affective upset is not to be confused with a chronic state of disorganization within the personality. A neurosis, to illustrate, may be regarded as a chronic disorganization within the subject. The neurosis reveals itself by various symptoms—tics,

amnesias, paralyses, sometimes repeated emotional outbreaks, and others. The emotional outbreak is not to be confused with the neurosis. The outbreak is a natural event, a process. The neurosis is an assumed enduring state or condition within the subject.

The distinction between an acute affective disturbance (which is coexistent with the manifest emotion) and a chronic state of disorganization within personality is not always clearly drawn by psychologists. In psychosomatic medicine, for example, the term *emotion* is used broadly to include any chronic mental disorganization as well as the acute affective disturbance. This confusion is to be regretted. When a student reads H. F. Dunbar's, *Emotions and Bodily Changes, a Survey of Literature on Psychosomatic Interrelationships,* he is at a loss to know exactly what constitutes an emotion. When emotion is equated with persistent disorganization within the personality as a whole the concept becomes so broad that it is virtually useless in psychology. It is significant, however, that even in the broad view of psychosomatic medicine emotion is conceived for the most part as a *disorganization* rather than as an organized response.

For the present writer emotion is defined as an acute affective disturbance, an event occurring here and now and revealed to the observer in different ways. An emotion is not a neurosis, nor a persistent state of conflict, nor an attitude, nor a motive, nor a habit structure. All of these are important concepts which are related to the concept of emotion, but they must be distinguished from emotion.

To clarify the picture the writer suggests that the term *emotion* be employed exclusively in the more restricted sense to designate an event or occurrence. The emotion is something that happens here and now, in the present tense. The term *affect* might be used within clinical and dynamic psychology in a broad generic sense to designate all persistent affective states of the individual such as persistent anxiety or depression or euphoria or excitement. An *affect,* then, is a persistent affective state of the individual.

THE DEFINITION OF EMOTION

What is an emotion? Our answer would be as follows: *An emotion is an acute affective disturbance within the individual as a whole, arising from the psychological situation, and manifest in conscious experience, behavior, and especially through bodily changes which are regulated by the autonomic nervous system.*

This definition of emotion rests upon a fundamental distinction between what is given as psychological datum and what is assumed or inferred as hypothesis. The data upon which any definition of emotion must rest are found in three areas of fact: (1) conscious feeling or experience, (2) emotional behavior as observed objectively in animals and men, (3) physiological changes, especially those regulated through the autonomic nervous system and the midbrain patterns of response, as observed in the laboratory.

The *real* emotion is *assumed* to exist within the psychological individual as a whole. The assumption is based upon facts within all three of the above areas. The experiential, behavioral, and physiological manifestations of emotion are regarded simply as manifestations of an assumed acute event within the subject.

Now when the writer defined an emotion as an *acute affective disturbance within the individual as a whole* he was interpreting the data in terms of an assumed individual. The emotion, so conceived, was distinguished from organized attitudes (of fear, of affection,

of hate, of disgust, etc.) and from organized motives (determinations to have revenge, to justify one's self, to woo and win, etc.).

There are two aspects of our definition which have been completely ignored by Leeper. We hope that the reason for this is that he accepts and agrees with them.

First, our definition states that *emotional disturbances originate within the psychological situation.* This part of the definition is important in distinguishing emotion from various illnesses (acute indigestion, headache) which are also disturbed affective states of the organism, and from appetites (hunger, thirst, fatigue) which originate from intra-organic conditions. The psychological situation, whether perceived or imagined, always involves an environmental component. Environmental events such as meeting a bear in the woods, being insulted by a rival, encountering one's love mate (to use William James' now trite illustrations) typically arouse emotions. Also seeing a movie or recalling a past experience may evoke emotional disturbance. Therefore, an environmental factor in the origin of emotion is what distinguishes emotion from various illnesses and from appetitive states.

Second, our definition affirms that *emotional disturbance involves bodily changes regulated through the autonomic nervous system.* The presence of marked changes in the activity of glands and smooth muscles is quite generally recognized as a sign of emotion. In fact, autonomic involvement is what distinguishes emotional from non-emotional disturbances. A distraction of attention can correctly be interpreted as a disturbed state, but unless the disturbance spreads to the involuntary nervous system and involves marked changes in the activity of glands and smooth muscles a simple distraction is not to be described as an emotion.

The above definition of emotion is assuredly imperfect. There is no single criterion which in itself is fully adequate to define emotion. But the three criteria of our definition, when they are met simultaneously, do indicate approximately what processes are and what are not to be regarded as emotions. This is about all one can expect of any arbitrary definition.

CONCLUSION

In conclusion, the problem of defining emotion in a positive and constructive way is still with us despite Professor Leeper's discussion. The following points have been made in the present analysis:

1. It is confusing and illogical to classify emotions into two groups, organized and disorganized. Organization and disorganization are logically opposed concepts and in strict logic only one of these concepts is required. Disorganization is the minimum of organization and *vice versa.*

2. Since emotion is an affective process, it must be defined in relation to other affective processes. The popular term *emotion* covers a wide variety of affective and motivational processes with little specificity.

To clarify the picture it is suggested that the term *affective process* be used broadly to include moods, sentiments, simple feelings of pleasantness and unpleasantness, interesting activities, and the like, and that *emotion* be employed specifically to designate that variety of affective process which is characterized by acute upset.

3. Of great practical importance in the definition of emotion is the distinction between a contemporary event revealed in conscious experience and behavior and bodily change, on the one

hand, and an assumed disorganization within the subject, on the other hand. Emotion as an event should be distinguished from neurosis, from attitude, from persistent motive, from persistent conflict, and from habit structure. The present writer's definition of emotion makes it possible to distinguish between emotion as a specific event and relatively stable affective states within the subject.

It is suggested that the term *affect* be employed broadly to refer to a persistent anxiety or depression or euphoria or excitement or other relatively permanent affective state within the individual and that the term *emotion* be reserved to designate a contemporary process or event.

4. To define emotion as an acute affective disorganization, etc., is entirely in agreement with the psychological facts. An acute upset, however, of itself is not a motive (since motives initiate, direct, and sustain action), but rather a condition which results from the dynamic interplay of motivating factors—from frustration or conflict or persisting pain or the sudden release of tension or the apprehension of some impending injury.

5. In a current paper based upon experimental results, the present writer has stressed the view that food-seeking drives are *organized* to preserve enjoyment and to relieve distress. Thus some affective processes are definitely organizing and the resulting behavior is organized. But despite this, it remains a fact of nature that acute affective disorganization occurs. It is quite proper to designate such disorganization as emotion.

REFERENCES

1. LEEPER, R. W. A motivational theory of emotion to replace "emotion as disorganized response." PSYCHOL. REV., 1948, **55**, 5–21.
2. YOUNG, P. T. *Emotion in man and animal, its nature and relation to attitude and motive.* New York: John Wiley and Sons, 1943. Pp. x + 422.
3. ——. Food-seeking drive, affective process, and learning. PSYCHOL. REV., 1949, **56**, 98–121.

[MS. received November 23, 1948]

THE PSYCHOLOGY OF INVENTION IN A VERY SIMPLE CASE

BY EDWARD L. THORNDIKE

Professor Emeritus, Columbia University

Fifty college graduates were asked to write meanings for the 30 words listed below. A group composed of 11 graduate students and 37 summer-school students had the same task for 22 of the 30: *abattoir, ambulatory, amerce, apogee, asseveration, auricular, besom, cacophony, cantilena, carmagnole, devolution, distrain, hermeneutic, largess, licentiate, murrain, nugatory, palindromic, peccancy, pediment, rhodolite, rune, saltatory, sedulous, sequestrate, sessile, shrievalty, tactility, teleological, umbel.* If the reader will take a few minutes to do the same for the first dozen or so of these words, the experience will be a useful introduction to what follows.

In the experiments these 30 words were mixed with novel words like *abbawl, accackle, debrag, illiberate,* and *perallot* formed with the prefixes *ab, ad, de, dis, in, per, pre, pro,* and *sub,* and novel compounds like *brush-mansion, desk-highland,* and *empire-fortress.* Since most of the 30 "real" words were also novel in the experience of most of the students, they could feel free to make inferences and guesses about any of them.

What happens is that the word and the task are responded to by one or more ideas until one is satisfactory enough to enable the person to write a meaning and go on to the next, or, very rarely, that nothing comes to mind which he is willing to put down as a meaning for the word.

In solving most of such problems most persons are easily satisfied. So *cacophony* is "phony cacao," "a substitute for cocoa," and "a substitute for cake." *Carmagnole* is "a car for magnolias," "a great car," "a highly decorated car," "red magnolia," and "by a magnolia tree." Very common are naive responses which hardly deserve to be called inventions, such as *abattoir,* "abettor," "ability to bat"; *ambulatory,* "related to the ambulance," "ambulance service," "conveyance like an ambulance"; *amerce,* "immerse," "to duck under water," "to plunge into water"; *apogee,* "apology," "a bogy"; *besom,* "be an amount," "be something," "bosom"; *hermeneutic,* "Hermes," "be in harmony with"; *rhodolite,* "road light," "after the manner of Rhodes"; *umbel,* "humble."

At the other extreme one finds ingenious inventions which would seem, to a jury which did not know that the correct meaning was correct, a more suitable meaning. For example, consider the following (the correct meanings are listed as a note on pages 193–4: *amerce,* "cover up" and "American Co."; *apogee,* "pogo stick"; *besom,* "a kind of ointment"; *hermeneutic,* "medicine for nerves"; *largess,* "a fat woman"; *murrain,* "torrent"; *palindromic,* "out of one's head"; *rune,* "crazy pig"; *umbel,* "shady spot."

Conceivably the causation of the response to any stimulus word in these experiments might include what the person had thought and done up to that point. Actually any influence from responses to words preceding the stimulus-word is extremely rare. In a sampling of over 1,500 responses I found no sure case and only three that were at all possible.

A college graduate might possibly

write any one of 20,000 or more meanings for any of these words. If he took meanings at random, he would be as likely to assign a certain meaning to any one unknown stimulus word as to any other; and such meanings as "a kind of animal," "a kind of plant," "a disease," "a medical term," "good," "bad," "large," "small," "red," "move in a certain way," and the like would be given often. His acceptance of such randomly chosen meanings would lighten his task in the experiment. Something like this happens occasionally in our subjects, but as a rule their responses are specific to the unknown stimulus word. Consider, for example, the 300 responses to *amerce, besom, murrain, rune, sessile* and *umbel.* The numbers of persons who gave no meaning rather than some pure guess were respectively 5, 6, 2, 6, 10 and 6 (35 in all). Of the 225 "wrong" meanings the only ones given for more than one stimulus-word were "plant" (for *besom* and *sessile*), "an object" (for *rune* and *umbel*), "a mineral" (for *besom* and *rune*), and "sad" (for *besom* and *murrain*).[1]

Similar evidence of specificity appears in the meanings given by a group of 47 students, about half of whom were college graduates, for the following artificial words: *hogance, sletter, slomp, trib, tropper,* and *vertous.* The only meanings given more than once were "fish" for *tropper* and *trib,* and "quick" for *sletter* and *trib.*

In all that follows we shall limit discussion to invented meanings, that is to say, responses uninfluenced by any knowledge of the correct meaning of the stimulus-word.

Note: For the reader's convenience I insert here the correct meaning for each of the words referred to in this article; and also the number (X) of invented meanings, (O) of failures to write any meaning, and (K) of responses influenced by more or less knowledge of the correct meaning. Since $X + O + K = 98$, these numbers are approximately percentages.

		X	O	K
abattoir	*n.* slaughterhouse. In percentages of 50 graduate students $X = 56$, $O = 10$, $K = 34$.			
ambulatory	*a.* 1. of walking. 2. able to walk. 3. adapted to walking. *n.* place for walking.	53	5	40
amerce	*v.* punish by a fine or penalty determined by the court.	83	15	0
apogee	*n.* 1. point in an orbit (especially of the moon) farthest from the earth. 2. highest or farthest point; acme.	63	19	16
asseveration	*n.* assertion. In percentages of 50 graduate students $X = 78$, $O = 0$, $K = 22$.			
auricular	*a.* 1. of the ears. 2. of hearing. 3. perceived by the ear. 4. of or pertaining to an auricle.	29	12	57
besom	*n.* 1. broom made of twigs. 2. broom.	72	21	5
cacophony	*n.* unpleasant sound; unpleasantness of sound.	59	25	14
cantilena	*n.* melody to be sung; graceful melody; a kind of medieval French song.	63	25	10
carmagnole	*n.* dance and/or song of the French Revolution.	54	36	8
devolution	*n.* 1. a devolving; act, process, or fact of passing on or down. 2. the opposite of evolution. In percentages of 50 graduate students $X = 72$, $O = 0$, $K = 28$.			

[1] Forty responses showed more or less knowledge of the word.

distrain	*v.* 1. in law, to seize or hold goods in order to satisfy a claim. 2. to levy a distress upon.	90	7	1
hermeneutic	*a.* that interprets or explains.	68	29	1
largess	*n.* 1. generous giving. 2. a generous gift.	60	15	23
licentiate	*n.* 1. holder of a license to practice an art or profession. 2. holder of a certain degree (lower than a doctorate) from a European university.	75	11	12
murrain	*n.* 1. a disease of cattle. 2. a plague.	59	21	18
nugatory	*a.* worthless; futile. In percentages of 50 graduate students $X = 74$, $O = 16$, $K = 10$.			
palindromic	*a.* that reads the same backward as forward. In percentages of 50 graduate students $X = 68$, $O = 26$, $K = 6$.			
peccancy	*n.* sinfulness; wrong doing. In percentages of 50 graduate students $X = 72$, $O = 20$, $K = 8$.			
pediment	*n.* low triangular gable.	80	4	14
rhodolite	*n.* a kind of red garnet.	62	16	20
rune	*n.* 1. one of the characters in an ancient alphabet of Scandinavian and Teutonic peoples. 2. something written or engraved in such characters. 3. verse; poem; song.	56	26	16
saltatory	*a.* 1. leaping; adapted for leaping. 2. progressing by leaps; not steady and gradual.	72	12	14
sedulous	*a.* diligent; painstaking. In percentages of 50 graduate students $X = 58$, $O = 10$, $K = 32$.			
sequestrate	*v.* 1. confiscate. 2. sequester.	34	24	40
sessile	*a.* 1. attached by the base instead of by a stem. 2. attached permanently.	60	30	8
shrievalty	*n.* the office, term of office, or jurisdiction of a sheriff.	65	27	6
tactility	*n.* fact, quality, or condition of being tactile. In percentages of 50 college graduates $X = 62$, $O = 4$, $K = 34$.			
teleological	*a.* of or pertaining to teleology; relating to final causes; pertaining to design or purpose in nature.	70	14	14
umbel	*n.* flower cluster in which stalks nearly equal in length spring from a common center and form a flat or slightly curved surface.	62	24	12

The stimulus word may act as a unitary total, or piecemeal by one or more of its parts of features. Thus *rhodolite* as a total may suggest and make acceptable "animal," "a chemical," "a flower," "a fluid," or "a kind of hard material"; *rhodolite* as a sequence of syllables may suggest and make acceptable "road light," "auxiliary lamp for road use," and "light like rhodora"; the first part of it may produce "after the manner of Rhodes" and "wild rhododendron"; the last syllable may produce "light," "luminous," and "bite."

In general the different parts or features of the stimulus words vary in the weight or potency they have for determining the response in different persons. For example, in the case of *hermeneutic*, "has to do with nerves," "pertains to the nerves," "neurological term" and "nervous" give evidence that *neu* had great weight; "hermetically sealed" and "hermetic" give evidence that *neu* had little weight. "Swift like Hermes" and "swift" give evidence that *herm* had great weight. "By oneself," "lonely," "live alone," and "of the nature of a 'hermit,'" give evidence

that *herme---t--* outweighted *neu------ic.* "Partly mechanical" and "in a half measure" give evidence that *he-m* was especially potent. In the case of *teleological, tele* and *logical* are about equally potent in the following: "long distance logic," "make logical by wire," "logical at a distance," and "logic of wiring." *Tele* is much more potent in "can be seen at a distance plainly," "far distant," "pertaining to telephones" and the like; *logical* is much more potent in "make logical," "observe logically," "logical writing," and "systematic."

The piecemeal activity of the stimulus word was especially common in these experiments because the words occurred among an equal number of tasks in getting meanings for artificial derivatives or compounds. This helped to set the mind toward the consideration of the parts of *abattoir, ambulatory, amerce,* etc. If the reader wrote meanings for a dozen of these words as suggested, his wrong meanings probably included fewer cases of piecemeal activity and more of activity as unitary totals than our records show.

The feature of a word which is potent in determining the response to it may be its pattern, especially the number of syllables it has and the place of the accent. The experiments afford one perfect illustration of this. When *rhodolite* is said to mean "the living of people in mountain caves," the causation is via *troglodite,* which has no syllable identical with *rhodolite,* but is very like it in pattern. Both words have three syllables, with the stress on the first, with ō, ŏ and ī as the vowel sounds, and with much more than chance similarity in the consonantal sounds, d, l, r, and t in the one, and d, g, l, r, t (twice) in the other.

An influence of the pattern of the stimulus word is, however, rare—very much rarer than Gestalt psychology should expect. I have examined all the responses to the 30 words in the list given above and also the responses to a dozen made-up words such as *accackle, debrag, hogance, irraly, perallot,* and *vertous.* The likeliest cases other than *rhodolite* are:

1. *ambulatory:* "place that babies are put to continue growth," ? via incubator.
2. *apogee:* "term used in literature," ? via trilogy.
3. *auricular:* "prophetic," ? via oracular.
4. *carmagnole:* "flower with many small, waxy, white petals," ? via camellia.
5. *carmagnole:* "Carmichael."
6. *largess:* "container," "container for food supplies," "storage vessel for food."
7. *licentiate:* "inebriate."
8. *murrain:* "earth," ? via terrain.
9. *palindromic:* "panorama."
10. *palindromic:* "many meters," ? via polymetric.
11. *perblink:* "bird," ? via bobolink.
12. *saltatory:* "dictatorial."
13. *shrievalty:* "frivolity."
14. *teleological:* "teaching," ? via pedagogical.
15. *teleological:* "utilitarian."

I may have missed some cases, but on the other hand some of these fifteen may not be genuine cases of the potency of "pattern." It seems unlikely that pattern is the determining factor in much more than one-half of one per cent of the wrong definitions.

The meaning given is, of course, often arrived at by one or more intermediate processes. So *hermeneutic,* "vacuum operated," probably was mediated by "hermetically sealed," "air tight," and "vacuum." In the case of *saltatory,* "oceanlike" was probably mediated by "salty," "beneficial" by "salutary," and "greeting" by "salutatory." In the case of *shrievalty,* "con-

fession" was probably mediated by "the act of shriving," "wrinkles," by "shrivelling up," and "feudal tenure" by "chivalry."

A response evoked in the mind or brain is a living thing, and may grow before it is put down on paper. So *hermeneutic* produces not only "sealed" but "sealed by air" and "seal by means of water." It produces not only "of the nature of a hermit" but also "measuring the depths of the hermit's shallow thoughts." So *apogee* produces not only "to excuse" but also "the written form of an excuse," and *saltatory* not only "salty" but also "salt of the earth," "seasoned," "spicy," and "sea-airish"; *murrain* produces not only "pertaining to rain" but also "hail or rain" and "the rain spoiled our party." And *sequestrate* produces not only "put aside" but also "put away for future reference."

It is reasonable to assume that most of the "wrong" responses given by the 98 persons might have been produced by one mind or brain, and that most of the connections causing any of them might have existed as potentialities of one mind or brain.[2] For example, all of the following could have been potentialities of one brain:

distrain as a total
- (1) "disdain"
- (2) "distrait"
- (3) "distrust"
- (4) "distend"
- (5) "restrain"

dis + *train* (of a railroad)
- (6) "derail"
- (7) "derail a train"
- (8) "jump the track"
- (9) "to put off the track"
- (10) "opposite of entrain"
- (11) "get off a train"
- (12) "alight"
- (13) "leave the train"
- (14) "miss a train"
- (15) "not to take a train"
- (16) "empty of passengers"
- (17) "to separate one train from another"
- (18) "take from train"
- (19) "unload from train"
- (20) "unload"
- (21) "remove"

dis + *train* (of practice, etc.)
- (22) "break a habit"
- (23) "make weaker"
- (24) "not to train"
- (25) "not trained"
- (26) "to train apart"
- (27) "to untrain"
- (28) "to remove training"
- (29) "lose the art of one's training"
- (30) "unlearn"
- (31) "those who might have failed to keep up with the training"

[2] Of course, if any such responses did occur, the possibilities for others would be changed.

dis + strain	(32) "avoid straining the human body" (33) "prevent a strain" (34) "relieve from strain" (35) "relieve the strain" (36) "release oneself from a strain" (37) "to relieve of uneasiness" (38) "surpass the limit of elasticity" (39) "strain away from" (40) "without strain"
di + strain	(41) "two strains"
− strain	(42) "a condition of strain"
	(43) "strain"
dis −	(44) "inharmonious"
	(45) "discord"

The assumption that most of what did happen in the 98 brains might have happened in one brain will help us to trace the causation of the various inventions of meanings. It also enables us to illustrate the important fact that two or more of the processes in a mind or brain may cooperate to determine the response. Thus we may account for *accackle* "berate" by *accackle* "make a noise" cooperating with *accackle* "attack." We may account for *auricular* "blood circulation" by *auricular* "circular" cooperating with *auricular* "pertaining to the heart." *Besom* "large and beastlike" clearly shows cooperation of *besom* "big" and *besom* "beast." So does *rhodolite* "light of plant," in which "flower," "plant family," etc. cooperate with "automatic light," "red light," "luminous," etc. So does *saltatory* "sarcastic speech," in which "statement" and "sharp" cooperate. ("Statement" and "sharp" are offspring of "salutation," "greeting," etc., and "astringent"—"much saltiness.") Why does *sequestrate* evoke "an orderly follower"? Probably because *sequestrate* evokes "to follow" and also "to do in turn." Why does *rune* evoke "waste land"? Perhaps because of *rune,* "sand dune," alone, but perhaps with cooperation from *rune,* "ruin."

We cannot expect to trace the causation of every response in these experiments. Some are due to peculiar events in a person's history; a few are, as has been noted, due to the search for any word that will serve the purpose. There is also a margin of unpredictability in such experiments because the events in a person's brain are not perfectly represented by what anybody knows about him, or what he knows about himself. But it is remarkable that one can find a reasonable causation for so many of them. Consider the responses to *amerce* by 83 persons (50 of the graduate group + 48 of the mixed group − 15 who wrote no meaning for it). None of the 83 gave the correct meaning, and probably none of them had ever encountered it. There were 71 different responses as follows: (1) against, (2) against traffic, (3) American Co., (4) amorous, lovely, (5) away from mercy, (6) baptism, (7) to be undeserving, (8) to be put under water, (9) bitter (two occurrences), (10) come above the surface of a liquid, (11) come together, (12) commerce, (13) consisting alone without intercourse or relations, (14) to cover up, (15) deprive, (16) disagreeable, (17) disband, (18) dissemble, (19) to distress (two occurrences), (20) divert trade, (21) draw from, (22) to duck under water, (23) emerse (two occurrences), (24) enter, (25) exclamation "a-mercy," (26) float, (27) forgive, (28) to gather together,

(29) go beneath, (30) go to water, (31) to go under completely, (32) to have mercy, (33) to hire (mercenary), (34) immerse (four occurrences), (35) to introduce, (36) join together, (37) keeping apace, (38) look, (39) make wealthy, (40) merciless, (41) merge, (42) move together, (43) new kind of water tank, (44) not trading with, (45) opposite of coerce, (46) persuade, (47) to plunge, (48) to plunge into water, (49) pull away, (50) to push up or out, (51) to put down, (52) put into, (53) to put under (three occurrences), (54) run away from, (55) show no mercy, (56) steeped in, (57) subdue (two occurrences), (58) submerge (three occurrences), (59) to take out of, (60) to take out of the water, (61) thank you, French, (62) to thank some one from French, (63) a thing that has been submerged, (64) trade, (65) to trade from, (66) travel, (67) to travel away from, (68) under force, (69) to urge, (70) weigh, (71) without mercy.

We can trace the causation of 60 of these with some surety.

amerce evokes 34, "immerse," 23, "emerse," 12, "commerce," 45, "opposite of coerce," 4, "amorous" (perhaps aided by a mispronunciation), 9, "bitter" via French *amer*, *amère*, and 16, "disagreeable" also via *amer*.

a + *merce* evokes 25, "a-mercy" (exclamation), 32, "have mercy," 71, "without mercy," 5, "away from mercy," and 55, "show no mercy."

a + *m* + *e* + *r* + *c* + *e* evokes 3, "American Co."

a evokes 1, "against."

merce evokes 41, "merge," 33, "to hire" (mercenary), and French *merci* or *remercie* producing 61, "thank you," and 62, "to thank some one."

If 34, 12, 41, 32, 5, and 1 come to mind or brain and evoke responses which are written, though they themselves are not, we have reasonable explanations of the following by second-order connections:

immerse evokes 22, "to duck under water," 48, to "plunge into water," 58, "submerge," and 6, "baptism."

immerse plus the *a* of the original stimulus evokes 10, "come above the surface of a liquid." Or perhaps the causation here is from *merge* via emerge.

commerce evokes 64, "trade" and 39, "to make wealthy."

merge evokes 69, "urge," 11, "come together," 28, "to gather together," 36, "join together," and 42, "move together."

merge plus the *a* of the original stimulus evokes 17, "disband."

have mercy evokes 27, "forgive."

away from mercy evokes 7, "to be undeserving."

against evokes 2, "against traffic," perhaps with some cooperation from commerce.

If 48, 59, the *water* of 22 and 48, and 64 come to mind or brain and evoke responses which are written though they themselves are not, we have reasonable explanations of the following by third-order connections.

to plunge into water evokes 47, "to plunge"; 8, "to be put under water"; and 52, "put into."

submerge evokes 63, "a thing that has been submerged"; 31, "to go under completely"; 29, "go beneath"; 53, "to put under"; and 14, "to cover up."

water evokes 30, "to go to water"; 26, "float"; and 43, "new kind of water tank."

trade evokes 65, "trade from"; 20, "divert trade"; and 44, "not trading with"—in all three cases probably with some cooperation from the *a* of the original stimulus.

As fourth-order connections we have:

put into evokes 56, "steeped in" with some cooperation from *submerge, water,* and *immerse;* and perhaps 35, "to introduce."
put into plus the *a* of the original evokes 59, "to take out of"; and 60, "take out of the water."
put under evokes 57, "subdue."
sub evokes 68, "under force."

As possible fifth-order connections we have:

subdue evokes 51, "put down."
take out of evokes 50, "push up or out"; 21, "draw from"; and 49, "pull away."
introduce evokes 24, "enter."

We are left with "consisting alone without intercourse or relations," "dark," "deprive," "dissemble," "distress," "persuade," "run away from," "travel," "to travel away from," and "weight."

Amerce is one of the words in the case of which I have had more than average success in tracing the causation of the "wrong" meanings. I had least success (fewer than 30 per cent accounted for) with *palindromic.*

In these experiments the process of invention is very simple because the stimulus word provides the essential data, and any meaning that satisfies the person can bring the search to an end. In most inventions a wider survey of data is required, and the criteria for accepting a solution are more exacting. Moreover, the piecemeal activity of a datum may be assisted by deliberate analysis, the assignment of weights may be much more complex, and the cooperation of events in the brain may be much more elaborate. But I doubt whether any activities radically different from those observed in our simple case are ever needed. It will be instructive to experiment with tasks of much wider scope or greater subtlety such as to invent ways and means to increase the reading of the Bible by college students, to reduce the divorce rate in the United States, or to discover how much more intelligent dogs are than sheep. I have done so, but with only one subject.

I append, somewhat irrelevantly, three notes on invention in general.

Psychologists have found by experiment that in certain cases successful invention is favored by discarding the tendencies which past experience and habits evoke. And Langmuir's shift from efforts to improve electric-light bulbs by getting a more perfect vacuum to filling the bulb chock full of some gas is a brilliant illustration of this. But on the whole the tendencies which past experience and habits evoke in the mind produce many more inventions than they prevent; and it is usually better to use them than to discard them. It is hard to know when to do the latter.

Cases are on record where dismissing a problem from thought causes it to be solved, the most famous being Poincaré's. This eminent mathematician would, after a certain amount of study of a problem, busy himself with other matters until some day the solution would suddenly come to him. Scientists in general are not so fortunate: if they don't find a solution when they are working on a problem, they rarely succeed by turning it over to unconscious cerebration.

Human inventions are notable examples of the fact that nature is creative as well as repetitive. Even in the brief and narrow experiment reported here we have persons thinking thoughts that nobody has ever thought before. We should, however, avoid the common mistake of supposing that creations, in the brain or elsewhere, cannot be as truly determined as the formation of water by the union of oxygen and hydrogen.

[MS. received December 9, 1948]

VALUE, NEED, AND OTHER FACTORS IN PERCEPTION

BY LAUNOR F. CARTER

University of Rochester

AND

KERMIT SCHOOLER

University of Michigan [1]

Recently social psychologists and those interested in personality theory have been urged by Bruner and his associates (1, 2, 3, 4, 5, 10, 11) to re-examine some of the basic phenomena in the field of perception. They have stressed the importance of need, personal value, stress, tension, etc. as organizing factors in the perception of stimuli ranging from clearly present objects to ambiguous material and on through imagined objects. Among the interesting results reported by these workers is an experiment by Bruner and Goodman (1) in which the importance of value systems in the perception of size is stressed. Using groups of poor and rich children they were able to show that not only did all the children greatly overestimate the size of coins, but that poor children overestimated coin size more than did rich children. Thus the child's need seemed to be associated with his perceptual judgment. These findings have been widely accepted as is evidenced by their almost immediate inclusion in Newcomb and Hartley's excellent *Readings in Social Psychology* (9) and in the new social psychology texts by Krech and Crutchfield (7) and by Sherif (12).

The influence of value on perception seems from these results to be so marked that poor children, to use an extreme example, perceive a quarter as being 52 per cent larger than it actually is. The need for valued objects is such for poor children that they not only perceive coins as larger than they actually are but they judge them to be even larger when the coin is present at the time of judgment than when the coin is absent and judgment is made from memory. Both because of the great importance these results would place on need and value systems in perception and because of their quick acceptance into the "facts" of psychology it was felt that the experiment would bear repetition.

Apparatus

The principal piece of apparatus used in this experiment is similar to that used in the Bruner and Goodman study. In one end of a 9″ × 7″ × 15″ metal box a 3″ circular hole was cut and covered by a disc of fine ground glass. Within the box and near the other end was placed a point source of light (Central Scientific Co. Concentrated Arc Lamp No. 87320) and its power unit. Between the glass plate and the point source a movable metal plate was inserted. A small hole was drilled in the center of the metal plate, forming a fixed aperture for the light to shine through. The plate was mounted on a rack and pinion gear in such a way that by turning a small knob on the side of the box, the plate with the aperture was moved forward or backward between the point source and the glass screen causing the disc of light falling on the ground glass screen to

[1] This experiment was completed while the junior author was at the University of Rochester.

be made smaller or larger. The intensity of the spot was measured to be 5.2 foot-lamberts. A millimeter scale superimposed on that portion of the rack which extended beyond the rear of the box made it possible to measure the diameter of the illuminated disc.

A series of thirty-six aluminum discs was also made. The diameters of the discs increased by increments of approximately 1 mm. from 12.5 mm. (½ inch) to 50 mm. (2 inches). Their thicknesses also increased with the diameter, the largest disc having a thickness of 3.1 mm. (⅛ inch). Five of these discs were of nearly the same diameter as the penny, nickel, dime, quarter and half dollar. A third part of the equipment consisted of a series of five cardboard discs (light gray in color), each one the same diameter as one of the above mentioned coins. Another part of the equipment consisted of these five coins themselves.

PROCEDURE AND SUBJECTS

The subjects for this study were 48 children whose ages ranged from nine to eleven years. Twenty-four of them have been designated as the "poor" group. They were obtained through the cooperation of a settlement house, and there can be little doubt that they are truly economically poor children. About one-third of these children were negro. The other twenty-four subjects have been designated as the "rich" group. They were all students of the fourth grade in a public school located in a well-to-do suburban area. It has been estimated by a responsible authority that the annual incomes of the parents of these children ranged between $5,000 and $35,000. Scores on an intelligence test indicate that all of the rich group were of normal intelligence; while no test scores were available for the poor group it is presumed that they were of normal intelligence since they were all attending the regular public or parochial schools.

Each subject, in his turn, was seated facing the light box in such a way that the glass screen was slightly below eye level, and the adjusting knob was easily reached. After demonstrating the apparatus to the child and allowing him to turn the knob by himself, the experimenter adjusted the spot of light to its largest position and requested the subject to make the spot of light as big as he believed a penny to be. In this manner the subject made ten adjustments, two for each coin denomination: one of these from the "out" position, that is, starting with the spot of light at its greatest diameter, and one from the "in" position, with the spot at its smallest diameter. After this task was completed, a penny was placed immediately to the side and in line with the front of the apparatus on a plain white sheet of paper in the horizontal plane, and the subject was again requested to make the spot of light as large as the coin before him. After the judgments for the penny were made, a nickel was placed on the paper and the penny removed; this was followed by the quarter, then the dime and finally the half-dollar. Again, ten judgments were made; two for each coin. Following this task, the subject was presented, in the same manner as for the coins, aluminum discs, and then cardboard discs approximately the size of a penny, a nickel, a quarter, etc. When each judgment was completed, the experimenter recorded its magnitude after reading the scale on the rear of the light box.

Individual aluminum discs from a part of the series of discs were next presented to the subjects in accordance with a predetermined order. The subjects were asked to decide whether the particular disc before them was larger or smaller than a 50 cent piece. Judgments of "the same" were not allowed.

After all the responses had been recorded, another part of the series of discs was presented, and the subject was asked to decide whether the disc before him was larger or smaller than a 25 cent piece. This procedure was repeated for comparisons with the dime, nickel, and penny.

RESULTS

Part I. Use of the method of average error. Let us first consider the obtained differences between socio-economic groups. We are led to believe by the results of the study by Bruner and Goodman that there will be a difference, that the poor children will exaggerate the size of the coins significantly more than the rich children, and that this difference is consistent with the hypothesis that "the greater the individual need for a socially valued object, the more marked will be the operation of behavioral determinants." Table I shows that in some cases a difference does exist between the groups, but not always in the same direction as would be expected according to the hypothesis stated above. An examination of the "t's" will show that the differences between the judgments of the rich and of the poor group are so small as to be insignificant in those cases where the coin is present, and also where judgments were made regarding the aluminum discs and the cardboard discs. When the data for memory are combined and subjected to analysis of

TABLE I

AVERAGE JUDGMENTS OF DISC DIAMETERS IN MILLIMETERS
(Number of subjects in each group is 24.)

	Dime		Penny		Nickel		Quarter		Half dollar	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Coin present										
True value	17.8		19.0		21.2		24.1		30.5	
Rich group	16.3	1.32	17.6	1.19	21.0	1.81	25.4	1.79	33.1	3.46
Poor group	16.5	1.51	18.6	1.84	21.2	2.01	25.7	2.17	33.9	3.58
$t_{rich-poor}$	+.51		+2.19*		+.33		+.48		+.77	
Aluminum discs										
True value	17.9		19.1		21.1		24.2		29.9	
Rich group	17.1	1.73	18.9	1.68	21.8	2.03	25.3	2.05	31.8	2.87
Poor group	16.9	1.40	19.0	1.59	21.4	1.97	24.9	1.83	31.2	1.27
$t_{rich-poor}$	−.70		+.42		−.67		−.43		−.73	
Cardboard discs										
True value	17.5		18.7		21.2		24.1		30.6	
Rich group	16.7	1.68	18.2	1.43	21.3	1.37	24.4	1.85	31.9	2.46
Poor group	16.4	1.30	17.8	1.59	20.9	1.73	24.3	1.98	31.9	2.95
$t_{rich-poor}$	−.18		−.90		−.87		−.68		.00	
From memory										
True value	17.8		19.0		21.2		24.1		30.5	
Rich group	16.1	2.18	18.0	1.86	20.5	2.14	26.1	2.48	33.3	4.11
Poor group	17.0	1.69	19.3	2.20	21.6	2.11	27.3	4.00	35.9	5.31
$t_{rich-poor}$	+1.22		+2.16*		+1.75		+1.56		+1.86	

* Significant at 5% level; + indicates poor group's judgment is larger, − indicates poor group's judgment is smaller.

variance the differences between the rich and poor groups are significant at past the 1 per cent level, although the differences for each separate denomination are not significant at this level. These results indicate that when judgments are being made with physical objects present as a standard of reference there are no significant differences in judgment between the rich and the poor subjects, but when the judgments are made from memory the poor children imagine the coins to be larger than do the rich subjects.

In this part of the experiment no statistically significant differences were found between rich and poor groups when real objects were present, but differences were found when judgments were made from memory. The evidence of exaggeration or of underestimation when judgments are compared with true values will be examined next. Once again we turn to the Bruner and Goodman study as a point of reference. In accordance with the first hypothesis of their study, that "the greater the social value of an object, the more will it be susceptible to organization by behavioral determinants," we should expect the effect for valued objects to be positive and proportional to the value or denomination of the coin. Figure 1, which can be obtained from Table I, shows the percentage exaggeration of each judgment compared with the true value of the disc being judged. In the figure we have combined the data for both the rich and poor groups for the coin-present situation, for the aluminum discs and for the cardboard discs since there were no significant differences between the two groups. The memory data for the rich and poor groups are shown separately. For ease of comparison we have also plotted some of Bruner and Goodman's data on this graph. The most striking aspect of these data is that we find not only some exaggeration, as Bruner and Goodman found, but also underestimation. Tests of significance show that, in general, the larger discs are significantly exaggerated and the smaller discs are significantly underestimated.[2] This generalization applies to cardboard and metal discs as well as to coins and memory of the size of coins.

Figure 1 also indicates the differences obtained between different types of discs. The question can be asked, for example, are the estimates found for the coin-present situation significantly different from the estimates for the cardboard discs? The answer given by Bruner and Goodman's study is "yes"; since the coins have value and the cardboard discs do not, the coins will be exaggerated more. We should also expect similar results when we compare coins with aluminum discs.

Inspection of Figure 1 shows that the only consistently outstanding difference is the tendency of the children in the poor group to make judgments from memory which are significantly larger than those made when a real stimulus object is present. Analysis of variance shows that in the case of the poor children the difference between the memory judgments and coin-present judgments is significant at past the .05 level of confidence. It will be noted that the judgments made from memory by the rich group fall very close to the judgments made while the coin was present. Thus it seems that the difference in the rich and poor group's judgments made from memory is due to the significant

[2] To conserve space the extensive tables showing the statistical tests will not be presented. The authors will be glad to supply copies of these tables on request. As a rough generalization any exaggeration or underestimation of 4 per cent or greater is significantly different than the true value at the 5 per cent level of significance and the corresponding figure for the .01 level of confidence is 5 per cent.

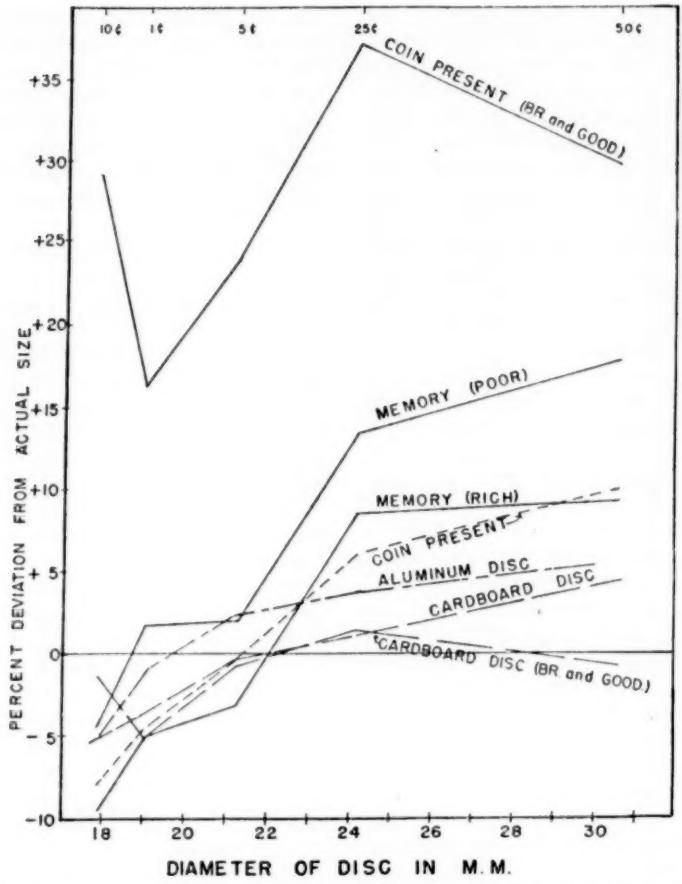

FIG. 1. Percentage underestimation and overestimation of the size of various discs.

overestimation of the poor group, while the rich group's judgments are similar to the coin-present judgments.

There is little difference in the accuracy of judgment for different kinds of discs of the smaller sizes; a few comparisons approach statistical significance but no general superiority is shown. For the quarter and half dollar size, however, the coin-present judgments and the rich group memory judgments are significantly larger than the judgments for the cardboard discs. The judgments for the aluminum discs lie between those for the coin-present situation and the cardboard discs with only the 50¢ coin present—aluminum disc difference being significant.

Part II. Use of the method of single stimuli. In this part of the experiment the subjects were required to state whether an aluminum disc was larger or smaller than a coin of specified denomination. (No actual coin was present for comparison purposes.) The point of subjective equality of rich and

poor groups for each denomination of coin was determined by finding for each subject the value of that disc which was judged to be just larger than the coin. The arithmetic mean of these "just larger" values was determined. Since the increment between coins in the series was uniform, the mean "just smaller" judgment may be taken as 1 mm. smaller than the mean "just larger" judgment, and the P.S.E. may be assumed to be halfway between the two.

The mean judgments in millimeters for the dime, penny, nickel, quarter and half dollar for the rich and for the poor groups, respectively, were 18.0 and 19.2; 18.4 and 19.3; 20.3 and 21.2; 29.5 and 30.9; 37.2 and 37.6. An analysis of variance shows the difference between the two groups to be significant at past the .01 level of confidence. Thus we see again that from memory the poor group imagines the size of coins to be larger than does the rich group. By this method we find a small exaggeration of the true size of the dime, a small underestimation for the penny and nickel and a very large exaggeration for the quarter and half dollar.

DISCUSSION

All of the above data lead us to the belief that the perceived size of a coin is closely related to its real size when judgments are being made regarding actually present objects. We would suggest that there is a constant error involved in making these size judgments such that small coins are understimated and large coins are overestimated in size. At the same time our data indicate that when judgments are being made from memory with no comparison object present, poor children's judgments are influenced by their need systems and they imagine coins to be larger than they really are. Rich children on the other hand, perhaps having less need for money and also presumably more familiar with coins than poor children, make judgments from memory which correspond closely with similar judgments made with the coins present.

Since these results are at variance with those obtained by Bruner and Goodman it is proper to make an attempt at explaining the discrepancies. While a definitive answer to the factors involved in the perception of coin size may demand the further collection of data it is the authors' belief that the perception of coin size is but slightly determined by value systems as long as judgments are made from clearly visible, present objects. It seems that not only our own results, but those of several other experimenters are consistent with this position. In 1898 Wolfe (14) published the results of a relevant series of observations. He asked school children and college students to draw circles the size of various coins and rectangles the size of several bills. His conclusions were, "Young children underestimate the size of coins and bills. Mature persons of intelligence overestimate the size of the silver dollar, half dollar and quarter dollar. All classes of persons underestimate the size of the dime, nickel and bill." It may be that the consistent underestimation of the young children is to be explained as a function of the small amount of money available to children of that day. But it is of interest that the results for eighth grade children and college students are very similar to our own.

Mansvelt (8), a Dutch worker, has published another relevant experiment. This work was done with brown and white cardboard discs, squares, and triangles. There were 19 figures for each form with the circles varying in size from a diameter of 26 mm. to 74 mm. The squares and triangles were of such a size that they could be inscribed in

circles of the above sizes. The subjects fixated a given figure for 5 seconds, then closed their eyes while a cardboard containing a number of similar forms differing from each other in size was placed in front of them. On opening their eyes they indicated which of these figures was of the same size as the original stimulus figure. Mansvelt found that the two smaller circles were correctly estimated and the larger circles were very slightly overestimated. The smaller squares and triangles were underestimated while the larger ones were overestimated. These results correspond closely to our own and resemble those of Wolfe. On the other hand it seems that Bruner and Goodman's results tend to be out of line with those obtained by these other workers.

One wonders why their results seem to be so different from those obtained by others. No basic or obvious errors in procedure are apparent and yet, as with almost any experiment, several questions regarding their work can be raised. The Bruner and Goodman data are based on three groups—a group of ten rich, ten poor, and ten control children. As a matter of experimental technique there is something to be said for letting each child serve as his own control, particularly where a small number of subjects is employed. The composition of their control group is not specified, presumably because the judgments with the cardboard discs were considered only random variations from true judgments of size. Inspection of Fig. 1, where we have plotted their data, shows these cardboard disc judgments to be fairly consistent with our own. It may be that had a larger number of subjects been used Bruner and Goodman would have found for the cardboard discs a constant error such as is suggested by our data.

A result which caused Bruner and Goodman some surprise was their finding that the poor group exaggerated the size of coins more when the coin was present than they did from memory. On the other hand the results for the rich group, with the exception of the 50¢ piece, showed no differences between memory and coin-present judgments. It would be our guess that these unexpected results for the poor group are to be explained as a function of the relatively small number of subjects employed.

A difference between our stimulus object and Bruner and Goodman's is that the intensity of their spot of light was 16 foot-lamberts while ours was 5.2 foot-lamberts. We should have preferred to use stronger illumination but the larger point sources are quite expensive and we believe that, if anything, the lower intensity of our image would cause exaggeration of judged size relative to Bruner and Goodman's image. Holway and Boring (6) have shown that a bright disc is judged to be larger than a less intensely illuminated disc. Accordingly if Bruner and Goodman's disc was set to exactly the same physical size as ours, theirs would be expected to appear larger. Since their disc appears the larger, subjects using it as a comparison with coins would, for the same size coin, set the disc physically smaller than they would on our apparatus. Thus, if anything, we should expect them to have found slightly less exaggeration than our results might show.

Not only is the judged size of coins a function of their physical size and the intensity of their illumination, it is also dependent on their shininess or dullness when under the same illumination and also on the angle of incidence of the illumination, as Weil (13) has shown.

While it is perhaps a minor point there is some criticism to be made of using an iris diaphragm to cast the

"circle" on the ground glass screen. As Bruner and Goodman point out such a "circle" is formed as a nine-sided figure with each side being a chord of a two inch circle. If such an iris diaphragm is closed down on a disc having a diameter of 18 mm., the diameter of a circle formed on the points of intersection of the sides of the diaphragm will be 5 per cent larger than the 18 mm. disc. Thus the children were faced with the task of approximating the diameter of a circular object, the coin, with a spot of light which was not circular. Bruner and Goodman do not believe this to be important and they are probably right, but it would have been nicer had the variable "circle" been a circle.

Finally, then, it is suggested that the judgment of the size of coins or discs when they are physically present is a function of many factors, including their physical size, their color or brightness, their illumination, the angle of incidence of the illumination, a constant error factor, and perhaps, even, the value attached to the disc. The present results suggest that the value of a disc or the need associated with it is important as a determiner of judged size only when the stimulus object is equivocal or absent, as when judgments are made from memory.

SUMMARY

Rich and poor children made judgments regarding the size of coins from memory, when the coins were present, from aluminum discs and from cardboard discs. The rich and poor children's judgments were essentially the same with the exception that the poor children overestimated the size of coins when judgments were made from memory. It was found that the size of small coins is generally underestimated and the size of large coins is overestimated. These results and those from other work raise doubt as to the general importance of value systems as organizing factors in the perception of clear, physically present objects. Apparently needs and values may play a role when the stimulus object is equivocal or not present as in the case of judgments made from memory.

REFERENCES

1. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
2. ——, & POSTMAN, L. Tension and tension-release as organizing factors in perception. *J. Personal.*, 1947, **15**, 300–308.
3. ————. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
4. ————. Emotional selectivity in perception and reaction. *J. Personal.*, 1947, **16**, 69–77.
5. HANFMANN, E., STEIN, M. I., & BRUNER, J. S. Personality factors in the temporal development of perceptual organization. *Amer. Psychol.*, 1947, **2**, 284–285.
6. HOLWAY, A. H., & BORING, E. G. The dependence of apparent visual size upon illumination. *Amer. J. Psychol.*, 1940, **53**, 587–589.
7. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
8. MANSVELT, E. Over het schatten der grotte van figuren van verschillenden vorm. *Meded. u. h. Psychol. Lab. d. Rijksuniv. t. Utrecht.*, 1928, **4**, II, 134–137.
9. NEWCOMB, T., & HARTLEY, E. *Readings in social psychology.* New York: Holt, 1947.
10. POSTMAN, L., & BRUNER, J. S. Perception under stress. PSYCHOL. REV., 1948, **55**, 314–324.
11. ————, & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **83**, 142–153.
12. SHERIF, M. *An outline of social psychology.* New York: Harpers, 1948.
13. WEIL, R. An illusion of size. *Nature*, Lond., 1945, **156**, 636–637.
14. WOLFE, H. K. Some judgments on the size of familiar objects. *Amer. J. Psychol.*, 1898, **9**, 137–166.

[MS. received December 9, 1948]

A THEORY RELATING MOMENTARY EFFECTIVE REACTION POTENTIAL TO RESPONSE LATENCY[1]

BY C. J. BURKE[2]

State University of Iowa

INTRODUCTION

In his book *Principles of Behavior* (**4**) Hull has set up a theoretical framework within which the behavior of organisms in certain simple learning situations can be explained. The theoretical structure is a pattern of mathematically defined relationships which permits the derivation of laws relating various observable aspects of the experimental situation.

In certain respects, the system is to be regarded more as programmatic than as final and precise. Several of the basic relationships have been hypothesized in the absence of adequate evidence as to their form (**1**, p. 214). This does not mean that the relationships have been set up blindly. Rather, they have been selected to reproduce certain known characteristics of data previously obtained, but such known characteristics are often not very precise indicators of the functions involved. A present important task, both experimental and theoretical, is to devise critical tests for these functions and to amend them wherever necessary.

Recently, Hull, Felsinger, Gladstone, and Yamaguchi (**1**, **3**, **5**) have employed response latency as a measure for the quantification of reaction potential. Latencies were obtained in an instrumental learning situation involving a bar-pressing response. For each of 59 rats a measure of response latency was secured on every trial during the course of learning and on a fairly large number of trials after the latency asymptote had been approached (**1**).

The values of reaction potential (E)[3] were obtained through use of the method of paired comparisons (**5**). The reasoning underlying the application of this method can be summarized as follows:

(1) The latency (t_i) of any given animal on any given trial can be compared with the latency (t_j) of that animal on any other trial. The proportion of cases among the 59 animals in which $t_i < t_j$ can be determined.

(2) Response latency is a monotonically decreasing function of momentary reaction potential ($\dot{E}$) (**4**, p. 345). This means that any pair of values of momentary reaction potential will have an order inverse to the order of the corresponding values of response latency. Thus, whenever, $t_i < t_j$, the relation $\dot{E}_i > \dot{E}_j$ must hold.[4] It follows that the data provide the proportion of cases in which $\dot{E}_i > \dot{E}_j$ for any pair of trials selected.

(3) According to the principle of oscillation (**4**, p. 319), values of $\dot{E}$ for a given organism are normally distributed. Since this is the distribution on which Thurstone's method of paired comparisons is based, it is assumed that the scale values yielded by the method of paired comparisons will correspond quite closely

[1] This paper was part of a Ph.D. thesis done under the direction of Dr. Kenneth W. Spence at the State University of Iowa.

[2] Now at Indiana University.

[3] The meaning of these terms from Hull's theory will be discussed more fully in a later section of this paper.

[4] Since all theorizing in the present paper is at the molar level with reference to a single stimulus aggregate and a single designated response, there can be no risk of misunderstanding if the subscripts "S" and "R" are dropped from Hull's notation.

with the quantitative definition of reaction potential introduced in the *Principles of Behavior* (4).[5]

From the experimental observations and the scaling procedure values of the median response latency (t_m) and of the reaction potential (E) are obtained for each value of the number of reinforcements (N). It is possible to make an empirical plot of either E or t_m as a function of N. By making use of the corresponding values of E and t_m it is also possible to plot either of these as a function of the other. Gladstone, Yamaguchi, Hull, and Felsinger (3) present these graphs. They find that the relations obtained can be well represented by the functions:

$$E = 3.55(1 - 10^{-.0305\dot{N}}),$$
$$t_m = 74.05\dot{N}^{-1.60} + 0.43,$$
$$E = 2.845t_m^{-.483} - 0.599,$$

where $\dot{N}$ is the median number of reinforcements from the beginning of learning. Accordingly, the general forms of the three equations above are taken as approximations to the desired functional relationships.

This approach of Hull and his co-workers is worthy of consideration in its own right. It represents, however, a new attack on the problem of quantifying reaction potential through the use of response latencies, and is not a direct outgrowth of the original ideas presented by Hull in his *Principles of Behavior* (4). First, the theory in Hull's book is written in terms of the individual organism; the data used in the Hull, Felsinger, Gladstone, and Yamaguchi studies result from the pooling of data from different organisms among which genuine individual differences must be presumed to exist. That the quantifications which they have achieved are not precisely what would be obtained from a group of identical organisms is expressly recognized as a possibility by these writers. They are careful to point out, for example, that their data are subject to two sources of variation, the oscillation of behavior within the organism and the distribution of individual differences among the organisms. Second, the form selected for the middle equation above is not consistent with the form of this function obtained by a rigorous deductive procedure from the relationships postulated in *Principles of Behavior* (4), as will be shown below.

In a sense, then, the problem of putting Hull's original postulates relative to response latency to experimental test under the conditions assumed in the writing of the theoretical structure remains. So far as this problem is concerned, the presence of individual differences makes direct application of the theory impossible when the data arise from a group of organisms.

The present study is concerned with the implications of Hull's earlier theory for the behavior of the individual organism. Although it is closely related to the more recent work done by Hull and his co-workers, the fact that it is concerned with the individual organism leads to certain differences.

DEFINITIONS AND BASIC RELATIONS

Two problems are involved in the kind of test of Hull's postulates that is proposed here, one theoretical, the other experimental. The theoretical problem is the deduction from the relationships presented in *Principles of Behavior* of relations among observable quantities which are amenable to experimental test. The experimental problem is the design of an experiment in which the results of the theoretical work can be tested. The present

[5] For a more detailed presentation of the above reasoning and an example of the actual working out of selected values, the reader is referred to Hull, Felsinger, Gladstone, and Yamaguchi (5).

paper is concerned only with the theoretical problem. The experimental problem, together with experimental findings, will be reported in a subsequent paper.

Since Hull's book constitutes the point of departure for this paper, it is convenient to begin the discussion with a brief recapitulation of the relevant postulates from Hull's system.

(a) *Reaction potential* (E): The reaction potential represents jointly the effects of habit strength and drive strength (**4**, p. 245). If conditions are such that drive is held constant, the reaction potential is a positive, negatively accelerated function of the number of reinforcements as given by the following equation:

$$E = E_0 - (M_E - E_0)(1 - e^{-iN}). \quad (1')$$

M_E is the product of the drive (D) and the maximum to which the habit growth curve can attain under the given experimental conditions (m). E_0 represents the reaction potential at the beginning of the experiment. This may be present due to the reaction potential associated with an innate receptor-effector connection or as a result of previous learning and generalization.

(b) *Effective reaction potential* ($\bar{E}$): Through the effective reaction potential, the effects of inhibition are taken into account. If I is the amount of inhibition present at any given time, the effective reaction potential is defined by the expression (**4**, p. 284):

$$\bar{E} = E - I.$$

If the experimental conditions are such that the amount of inhibition developed is negligible, I may be assumed to be approximately equal to zero so that:

$$\bar{E} = E,$$

or, from (1'):

$$\bar{E} = E_0 - (M_E - E_0)(1 - e^{-iN}). \quad (1)$$

Equation (1) is one of the basic equations upon which the subsequent development is based. Its properties can be clearly seen from an inspection of Fig. 1.[6] It is seen that $\bar{E}$ grows from a value of E_0 at $N = 0$ at a gradually diminishing rate so that the value M_E is approached asymptotically as N increases without limit.

The validity of the application of equation (1) to any set of data depends, of course, upon the degree to which the assumptions of constant drive and negligible inhibition are fulfilled in the experimental situation which gives rise to the data.

(c) *Momentary effective reaction potential* ($\mathring{\bar{E}}$): Due to the influence of variables which cannot at present be measured or controlled, there is variability from moment to moment in the amount of effective reaction potential. As the effective reaction potential oscillates, the amount present at any moment is called the momentary effective reaction potential. Values of the momentary effective reaction potential are assumed to be normally distributed with variance σ^2 and mean $\bar{E} - 2.5\sigma$[7] (**4**, p. 319). At the asymptote, where $\bar{E} = M_E$, the mean is $M_E - 2.5\sigma$. The variance of the oscillation function is assumed to be independent of the number of reinforcements. In Fig. 1 the normal curves showing the oscillation of $\mathring{\bar{E}}$ at two selected values of N are presented.

(d) *Threshold* (L): Whenever the momentary effective reaction potential falls below a certain critical value, no response will occur. This critical value (L) is called the reaction threshold (**4**, p. 344) and is shown schematically on Fig. 1. According to the properties of the oscillation function, the probability that a given response will be elicited in some selected interval of time (provided, of course, that the time interval selected is somewhat longer than the physiological minimum required for the occurrence of the re-

[6] No importance should be attached to the particular values of the parameters which occur, as these have been chosen arbitrarily.

[7] The 2.5σ enters because Hull conceives of oscillation as mainly inhibitory. Thus the value of $\bar{E}$ is taken well above the mean of the distribution. The form of the laws would not be changed if it were taken at the mean.

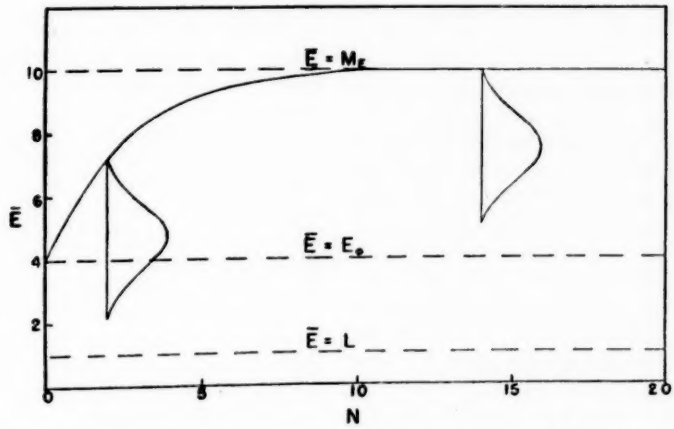

FIG. 1. A schematic representation of the relation between effective reaction potential ($\bar{E}$) and number of reinforcements (N).

sponse) is determined by the position of the threshold, the mean value of the oscillation function, and the oscillation sigma (**4**, p. 385). As learning proceeds, this probability increases until, as the limit of learning is approached, the probability is usually very close to unity.

(e) *Response latency* (t): Hull has postulated the following relationship between response latency and momentary effective reaction potential (**4**, p. 345):

$$t = \frac{a}{\dot{\bar{E}}^b},$$

where a and b are to be estimated from the data.

This function was slightly modified in the present study. The equation used is:

$$t - t_0 = \frac{a}{\dot{\bar{E}} - L} \qquad (2)$$

where t_0 is defined as the physiological limit of response latency for the particular organism at the time of occurrence of the response. Equation (2) is basic to the subsequent theory. It is represented graphically in Fig. 2, where the properties which led to the choice of equation (2) are exhibited. t is a monotonically decreasing function of $\dot{\bar{E}}$. As $\dot{\bar{E}}$ approaches the threshold L, the response latency increases without limit; the infinite latency which occurs when $\dot{\bar{E}} = L$ expresses the notion that a response will not occur unless the momentary effective reaction potential is above threshold. As $\bar{E}$ increases without limit, t approaches more and more closely to the value t_0. These features are apparent in Fig. 2. The constant b occurring in Hull's equation was tentatively taken as unity in the belief that the expression was sufficiently general to account for the experimental results without the introduction of the extra parameter. The units in which t and $\dot{\bar{E}}$ are expressed affect the value of the constant a.[8]

(f) *Behavior populations:* The principle of oscillation implies an inability to make an accurate prediction of the precise characteristics of a response in a specific situation. The latency of

[8] The form of the equations which apply to measurable quantities as given in the mathematical notes is of some interest in connection with the question of units. The quantity a occurs only in the ratio a/σ, which is to be evaluated from the data. Wherever $\bar{E}$, $\dot{\bar{E}}$, $\bar{E}_0$, M_E, and σ occur, they occur in dimensionless ratios. This means that it is immaterial what particular units of reaction potential are chosen. As the units of t are changed, the value of a/σ will change proportionately.

any given response occurrence depends, in the manner illustrated in Fig. 2, upon the value of $\dot{\bar{E}}$ which characterizes the state of the organism at the moment of the response occurrence. To predict the latency before the response had occurred would require a prior knowledge of the value of $\dot{\bar{E}}$. According to the principle of oscillation, such prior knowledge is impossible. At best, one can only know, from the form and location of the oscillation curve, the probability that the value of $\dot{\bar{E}}$ will lie between any two selected values. It is convenient to regard the oscillation curve as defining a population of potential behavior for the organism. The term "population" is used in the ordinary statistical sense. Just as the numbers corresponding to the heights of all adult males in the United States constitute a statistical population from which samples of numbers may be drawn, the numbers corresponding to momentary reaction potential constitute a statistical population from which samples of numbers may be drawn. The position of the oscillation curve changes during the course of learning. A number which is obtained upon a single occurrence of the response may be thought of as a sample of a single item from the population which is present at the time of the response occurrence.

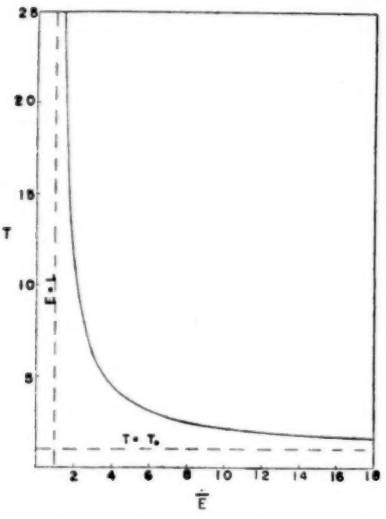

FIG. 2. A schematic representation of the relation between response latency (T) and momentary effective reaction potential ($\dot{\bar{E}}$).

It can be seen from an inspection of Fig. 1 that, when the value of $\dot{\bar{E}}$ is near the asymptote, changes in the position of the oscillation curve are small. This contrasts markedly with what is observed on the early portion of the curve where each reinforcement leads to a considerable change in the position of the oscillation curve. Because of the contrast two aspects of the learning process will be distinguished and designated as the *learning phase* and *the asymptotic phase.* In the learning phase each reinforcement brings about an appreciable increment in habit strength, and hence in excitatory strength, with consequent systematic behavior changes: during this phase samples of more than one item cannot be obtained from any behavior population. In the asymptotic phase the effect of each reinforcement is negligible and larger samples can be obtained, provided that the variables, other than the practice, which effect the placement of the oscillation curve, can be controlled.

The above discussion has been largely in terms of intervening variables and, consequently, non-operational. It is recognized, of course, that values of $\dot{\bar{E}}$ cannot be drawn directly from a population of such values. In the following section, the preceding discussion will be related to the problem of drawing samples of response latencies, which are easily amenable to

operational definition, from the corresponding behavior populations.

THEORY

In this section the implications of the postulates will be explored. A considerable number of testable conclusions, relating to the behavior of an individual organism as exhibited in the latencies of its responses in a given instrumental learning situation, will be derived. The derivations in this section will be presented graphically. The corresponding rigorous derivations are given in the mathematical notes which follow the paper.

The learning phase and the asymptotic phase will be discussed separately. Although the learning phase is prior to the asymptotic phase in the learning process, theoretical discussion is facilitated if the asymptotic phase is considered first.

A. The asymptotic phase. Equation (2) states that the response latency is a monotonic function of the momentary effective reaction potential. Thus, there is a unique, one-to-one correspondence between the values of $\dot{\bar{E}}$ and the values of t. It follows that the values of t must exhibit a variability related in some manner to the postulated oscillation of $\dot{\bar{E}}$. From these considerations the problem of determining the *variability function* of t arises.

The variability function of t can be derived from the oscillation function of $\dot{\bar{E}}$ and the relationship between t and $\dot{\bar{E}}$. The mathematical derivation is given in the mathematical notes. The graphical derivation is based on Figs. 3 and 4.

Figure 3 is a composite of three graphs. In the upper right section the relation between t and $\dot{\bar{E}}$ is shown. This portion of the graph is merely a part of Fig. 2. The lower right section contains the oscillation curve of

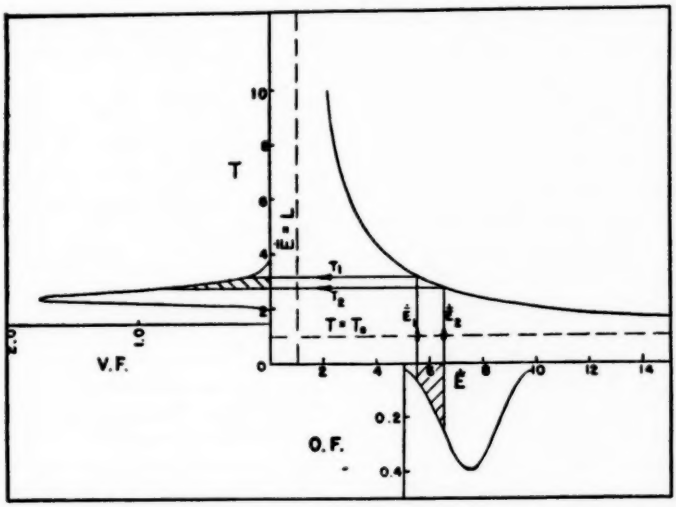

FIG. 3. A schematic representation of the relation between the non-cumulative variability function (V.F.) of T and the non-cumulative oscillation function (O.F.) of $\dot{\bar{E}}$ after 14 reinforcements as defined by the relation between T and $\dot{\bar{E}}$.

$\hat{E}$ with the axis of $\hat{E}$ common to this graph and the one above it. The oscillation curve has been drawn with the mean value at 7.5 units and a sigma of one unit. These values correspond to the values which characterize the oscillation function drawn for $N = 14$ on Fig. 1. Hence the oscillation function of Fig. 3 may be considered to represent the asymptotic phase in Fig. 1. The cross-hatched area in the lower right section represents the probability at any moment as defined by the postulated oscillation function, that $\hat{E}$ lies between two selected values, $\hat{E}_1$ and $\hat{E}_2$. The verticals corresponding to the two selected values of $\hat{E}$ have been constructed on the graph.

In the upper left section the variability function for t is presented. The axis of t is common to this and the graph on the upper right. The variability curve for t has been drawn so that the probability that t lies between any two selected values is equal to the probability that $\hat{E}$ lies between the corresponding values defined by the relation between t and $\hat{E}$. The horizontals labelled t_1 and t_2 correspond respectively to the values $\hat{E}_1$ and $\hat{E}_2$. Hence the probability that t lies between t_1 and t_2 must be equal to the probability that $\hat{E}$ lies between $\hat{E}_1$ and $\hat{E}_2$. This means that the size of the cross-hatched area for t relative to the total area between the variability curve and the t-axis must equal the size of the cross-hatched area for $\hat{E}$ relative to the total area between the oscillation curve and the $\hat{E}$ axis. Since the absolute size of either cross-hatched area depends upon the units arbitrarily selected in plotting the curve, it does not follow that the two cross-hatched areas drawn on Fig. 3 should be of equal size. The variability curve of t is so drawn that these relative areas will be equal no matter which two values of $\hat{E}$ are taken as $\hat{E}_1$ and $\hat{E}_2$. The analytical expression for this curve is equation (8) of the mathematical notes. A graphical method of constructing it will be given in connection with Fig. 4.

The information given in Fig. 4 is identical with that in Fig. 3; the only difference is that the cumulative form of the oscillation and variability curves has been used in place of the non-cumulative form. The normal probability function in Fig. 3 has been replaced with the probability integral and the variability function has been replaced with its integral. These cumulative curves have been introduced because they provide a more natural basis for deriving the variability function for t. This will be illustrated by the ease with which the cumulative form of the variability curve can be constructed.

The cumulative oscillation curve in the lower right of Fig. 4 has a value of 0.85 at the value 8.50 for $\hat{E}$. This means that 85 per cent of the values of $\hat{E}$ are, under the assumed conditions, less than 8.50 units. If the vertical corresponding to this value of $\hat{E}$ is followed up to its intersection with the curve in the upper right section, the corresponding value of t, represented by a horizontal on the graph, is found to be 2.33 units. Since the relationship between t and $\hat{E}$ is decreasing, a value of t will be larger than 2.33 units whenever the corresponding value of $\hat{E}$ is smaller than 8.50 units. Hence, values of t larger than 2.33 units will be found on 85 per cent of the trials under the conditions for which Fig. 4 has been drawn. It follows that values smaller than 2.33 units will be found on 15 per cent of the trials. Accordingly, the cumu-

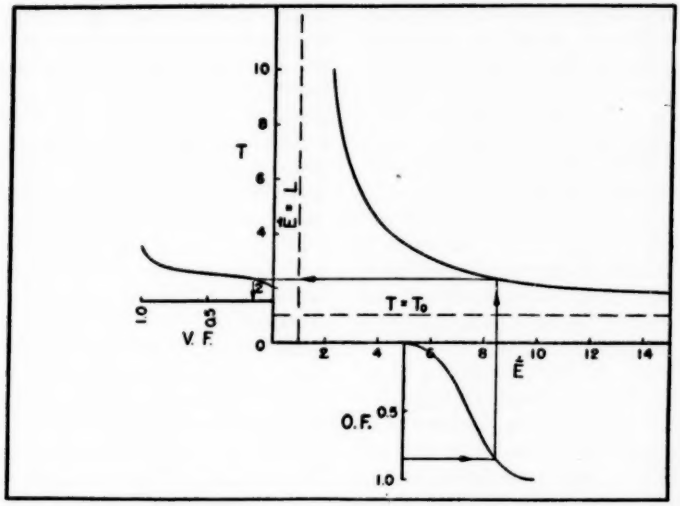

FIG. 4. A schematic representation of the derivation of the cumulative variability function (V.F.) of T from the cumulative oscillation function (O.F.) of $\hat{E}$ after 14 reinforcements and the relation between T and $\hat{E}$.

lative variability curve for t has the value 0.15 when t equals 2.33.

The operations described in the paragraph above can be repeated for other points until a sufficient number are obtained to permit the drawing of the curve. In this way the cumulative variability curve can be constructed. The expression for this curve is given in equation (12) of the mathematical notes. The non-cumulative variability curve can, in principle at least, be obtained directly from the cumulative form. The slope of the cumulative variability curve can be measured at as many points as is desirable. The graph of these slopes, as a function of t, is the non-cumulative variability curve.

The consideration of a hypothetical large group of identical organisms may facilitate understanding of the variability function. The conditions of a learning situation may be assumed to lead to a value of $\bar{E}$ which grows according to the curve in Fig. 1 for such a group. On each trial a response latency can be measured for each animal in the group. The distribution of response latencies that would be observed on the 15th trial (after 14 reinforcements) is described by the curves in the upper left of Figs. 3 and 4. The response latency distribution for a single organism which is at the asymptote of habit strength would also be described by these curves provided that other relevant experimental conditions are carefully controlled so that the variability in response latency from trial to trial can be entirely attributed to behavioral oscillation. In this way, a rather concrete illustration of the notion of a behavior population can be given.

Certain features of the variability curves can be observed on Figures 3 and 4: (1) They are positively skewed. (2) The median (or any other percentile) value can be determined. (3)

The physiological limit of response latency is not closely approached. The value of t_0 which has been chosen for these schematic diagrams is unity, as shown on Fig. 2. From the variability curves on Figs. 3 and 4, it can be seen that very few response latencies smaller than two units occur. The behavioral interpretation of this discrepancy is simply that the value of the drive which has been arbitrarily chosen for these computations is too small to give rise to a maximal effort on the part of the organism.

A feature of the variability functions which is not apparent from the diagrams, but which is demonstrated in the mathematical notes, is that the mean and standard deviation of this distribution are infinite. The occurrence of these infinite moments may or may not represent the behavioral facts. In this case, the result is due to the mathematical model, extrapolated beyond the range of the data. Other models may exist which resemble the present one over the range for which data have been obtained but have finite moments for the variability function. The writer is inclined to believe that the infinite moments possibly do have behavioral meaning. As evidence of a very large variance, the great variability of sample means relative to that of sample medians can be cited (1). The fact that the population mean fails to exist should not be construed as a prediction that an infinite mean will necessarily be observed in any finite number of samples. It does mean, however, that the possibility of a complete response failure cannot be completely rejected, even though the probability is very small. Any further discussion of this point would be pure speculation since the criterion for a complete response failure in an instrumental learning situation has not been adequately defined.

B. The learning phase. When a

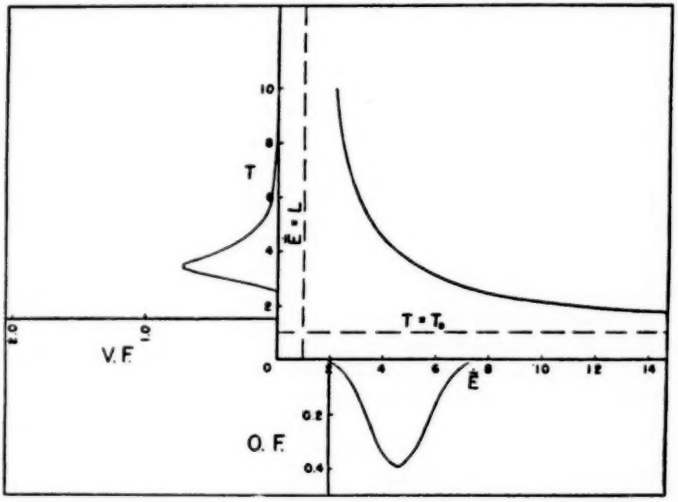

FIG. 5. A schematic representation of the relation between the non-cumulative variability function (V.F.) of T and the oscillation function (O.F.) of $\hat{E}$ after two reinforcements as defined by the relation between T and $\hat{E}$.

group of identical organisms is considered, the possibility of dealing with the variability function for t, even when the animals are not at the learning asymptote, becomes apparent. The observed latencies from such a large hypothetical group could be considered to define a behavior population of response latencies for each value of N. Since it is not possible to obtain even two identical organisms, the largest sample which can be obtained from each such behavior population is of unit size. Yet it is possible to derive, from a consideration of the behavior populations, certain features of the behavior of the individual organism which are amenable to experimental test.

In Fig. 5 the variability curves for a selected, non-asymptotic case ($N = 2$ in Fig. 1) are presented. Figure 5 is in all respects similar to Fig. 3, save that the mean value of the oscillation function is at 4.75 units instead of 7.5 units. This means of course that a different section of the curve relating t and $\dot{E}$ is used in the upper right. Aside from this, the method of constructing the variability curve in the upper left of Fig. 5 is identical with the method of constructing the variability curve in Fig. 3. The variability curve in Fig. 5 is strikingly different from the variability curve in Fig. 3, and represents the distribution of response latencies which would be observed in the group of identical organisms on the third trial. This curve cannot be determined empirically by a set of measurements upon a single organism since, in the early stages of the learning process, each reinforcement has a systematic effect upon behavior. A comparison of the variability curves in Figs. 3 and 5 reveals that: (1) Response latencies tend to be larger, on the whole, in the early stages of the learning process than in the later; (2) the variability in response latency is markedly greater in the earlier stages of the process; and (3) the skewness of the distribution is more marked in the earlier stages. Each of these variables—central tendency, variability, and skewness—can be studied as a function of the number of reinforcements. Again, the mathematical argument is given in the mathematical notes; the outline of the corresponding graphical argument is given in the following paragraphs.

TABLE I
THE VALUES OF CERTAIN PARAMETERS IN THE VARIABILITY FUNCTION OF t AT TWO POINTS IN THE LEARNING PROCESS

	Number of reinforcements	
	2	14
Mode	3.39	2.47
Median	3.71	2.54
Interquartile Range	0.96	0.26

A difference in the variability functions for t corresponding to a difference in the stage of learning is demonstrated in Figs. 3 and 5. A more detailed study of certain features of the two variability curves will be of interest. In Table I are given the mode, median, and interquartile range for the variability curves in Fig. 3 and Fig. 5, which correspond respectively to the oscillation curves in Fig. 1 for 14 and two reinforcements. The two values for the mode can be directly read from the variability curves in Figs. 3 and 5. For N equals 14, the values of the median and interquartile range can be read from the cumulative variability curve in Fig. 4; the median and interquartile range for N equals two cannot be read from any of the graphs presented in this paper.

The relatively small values of the median and mode after 14 reinforcements as compared with the values after two reinforcements indicate that response latencies tend to be lower at

the asymptote of learning than during the learning phase. It is possible, of course, to construct a variability curve of response latency for every value of N, to determine the median and mode of each such variability curve, and from these to construct the functions relating median response latency and modal response latency to the number of reinforcements. The analytical derivation of these functions is given in the mathematical notes. The curves representing the functions are given in Fig. 6. The solid curve is for the medians and the dashed one for the modes. Both curves exhibit the same general features, a relatively high value in the early stages, a gradual decrease as learning progresses, and an asymptotic approach to a value in excess of the physiological limit. As mentioned above, the fact that the physiological limit is not approached indicates that the drive strength which has been selected in this hypothetical case is not large enough to bring about a maximal effort on the part of the organism. The curve for the modes lies below the curve for the medians throughout and approaches a limit which is noticeably smaller than the limit approached by the curve for the medians. This indicates a positive skewness throughout the course of learning. A further theoretical prediction, indicated by the closer and closer approach of the curve for medians and the curve for modes as N increases, is that the skewness becomes less marked as learning progresses.

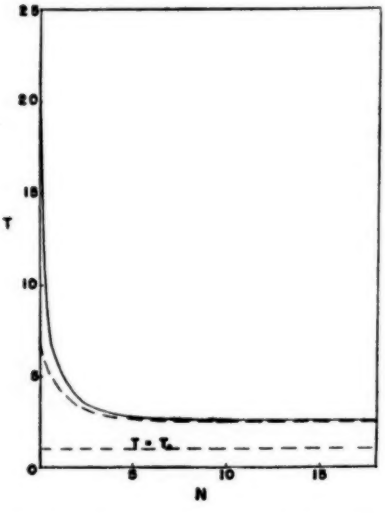

FIG. 6. A schematic representation of the function relating median latency (solid curve) and modal latency (dashed curve) to number of reinforcements (N).

From Table I it can be seen that the interquartile range is smaller at the asymptote than early in the learning. As with the median, the function relating the interquartile range to N can be determined easily when the variability curve has been constructed for each value of N. This curve has not been plotted, but shows essentially the same features as the curves in Fig. 6, a gradual decrease from an initially high value to approach asymptotically a value of about 0.25 units. Under an ideally maximal drive strength, this quantity would approach a value of zero. The prediction that arises from this function is that variability in response latency decreases as learning progresses, despite the fact that the oscillation sigma of $\hat{E}$ has been assumed to be independent of N.

C. Discussion. The theory relating response latency to the system presented by Hull (**4**) has been developed and certain testable deductions have been made from it. The possibility of dealing theoretically with a set of measurements of response latency obtained from a single animal at its learning asymptote has been demonstrated. A theoretical picture of the change in the median latency during the early stages of learning has

been established so that the possibility of dealing with data on individual animals in the early stages can also be considered. If the results reported by Felsinger, Gladstone, Yamaguchi, and Hull (1) can be taken as representative of latency data in well-controlled experiments, it is seen that the characteristics of response latency derived above are in rough, qualitative agreement with observations. A more precise, quantitative check on the theory is desirable nevertheless.

Certain assumptions have been necessary to the development of the theory. If the theory is to provide an adequate description of the data, the experimental procedure must provide controls which insure that these assumptions are met. In the following paragraphs the problem of designing an experimental procedure which will embody the necessary controls and thus provide a fair test of the theory is briefly discussed.

For the asymptotic phase, a distribution curve has been derived which represents the behavior population of response latencies for a given organism in the asymptotic phase. The experimental problem of the asymptotic phase is to obtain a sample from this behavior population which is sufficiently large to be a reasonably good approximation to the population itself. From such a large sample an empirical variability curve for response latency can be constructed to check the adequacy of the theoretical variability curve for the representation of the data. If the theory is to apply to the experimental results, conditions must be so arranged that all variability in response latency can be attributed to oscillation alone. If it is possible to make a large number of measurements of the latency of some instrumental response for an individual organism with all factors which systematically effect response latency controlled, the theory for the asymptotic phase can be tested.

During the learning phase the experimental problem is to obtain a sample of a single item from each of the behavior populations which occurs during the learning process. In other words, a measure of response latency is to be secured on each trial in the learning process under experimental conditions which control all variables that might effect response latency, aside from the effects of the two sources of variation under study—oscillation and number of reinforcements. In the early stages of the learning process, competing responses are quite likely to occur and the occurrence of such competing responses represents a source of contamination in the present test; hence it is desirable to select a response which is not particularly subject to the influence of competing responses. When such a series of response latencies has been obtained, it can be analyzed to determine whether the measures of central tendency, variability, and skewness for the latency distributions behave in the manner predicted by theory, aside from the effects of sampling fluctuation.

In the preceding paragraphs a tacit assumption that the necessary variables can be controlled has been made. If the variables are enumerated and the methods of controlling them are compared, certain contradictions appear. These contradictions have considerable influence on the experimental design. The next few paragraphs are devoted to a brief analysis of the situation which occurs when an attempt is made to control the major variables thought to influence response latency. These major variables are age, drive strength, amount of inhibition, warm-up effect, and external stimulation.

Obviously, the effects of changing age cannot be controlled, but they can be minimized by confining the entire experiment to a sufficiently short length of time. The size of sample required and the restrictions imposed by the necessity for controlling other variables will set a lower limit for the time over which the experiment extends. A compromise in which the various relevant variables are considered is necessary. In such a compromise, the necessity for keeping age changes small requires, at the very least, that there be more than one trial per day if a sample of sufficient size is to be obtained.

It is possible for drive strength to change either from day to day or within the course of the daily run. If hunger is used as the drive, the day to day variations are controlled by running under the same conditions of food privation on each day. Within the daily run two factors which might change the drive strength operate. The reinforcement following each trial may decrease the drive strength; the mere passage of time without appreciable quantities of food may increase it. The first effect can be minimized by using small amounts of reward and the second by confining the daily run to a short period of time.

The amount of inhibition should be negligible. In Hull's theory inhibition is primarily a work phenomenon which dissipates with time (**4**, p. 300). The amount of inhibition can be kept small by using distributed trials within the daily run.

A warm-up effect may be present so that the early trials within the daily run will have larger latencies than the later trials. If this is true, it will be necessary to discard one or more of the trials on each day in order to secure samples which are independent of the effect.

The means of controlling these variables are incompatible. For age changes, the number of trials per day should be large. To maintain drive strength at a constant level, the total time interval within which the daily run occurs must be small. To keep the inhibition negligible, the trials must be well spaced. Moreover, the possibility of warm-up effect may necessitate the discarding of some of the measures. In spite of these contradictions, a fairly large sample must be obtained with effective control of the variables.

An experimental test of the present theory has been made and will be reported soon. In that test, a compromise procedure was adopted for the controls with no assurance that it would prove effective. Consequently, a test of the results was made prior to the theoretical analysis. If the controls are effective, all measures secured during the asymptotic phase arise from the same population. Under these conditions both the day-to-day variability and the variability from position to position in the daily run can be accounted for in terms of the basic oscillation.[9] To determine whether this is true for any given animal the techniques of analysis of variance can be used. If, for any animal, either the day-to-day variability or the position-to-position variability is significantly larger than the basic oscillation, as evaluated by the interaction variance, the data from that animal are simply irrelevant as a test of the present theory. When the controls have been ineffective, the assumptions on which the theory is based are not fulfilled. In this connection, it should be noted that the basis for selection of data outlined

[9] The interaction variance, positions by days, would be taken as an estimate of the basic oscillation.

above is independent of, and logically prior to, the particular theory which is to be tested.

SUMMARY AND CONCLUSIONS

A rather comprehensive theory for measurements of response latency obtained from an individual organism in an instrumental learning situation has been developed from relationships postulated by Hull in his *Principles of Behavior* (**4**). The testable relationships which have been derived are as follows:

(a) The frequency distribution of response latencies when all variables known to effect response latency, including practice, are controlled so that the latency measurements are essentially a random variable: A precise function which purports to describe this distribution has been derived and is available for fitting to observed frequency distributions.

(b) The median latency as a function of number of reinforcements: Again a precise function is available for purposes of curve-fitting. According to this function the median latency should decrease rapidly in the early stages of the learning process, less rapidly as learning proceeds, and approach an asymptote as the number of reinforcements increases without limit.

(c) Variability of response latency: It is predicted that the variability of response latency decreases as the number of reinforcements increases, approaching a limit asymptotically as the number of reinforcements increases without limit.

(d) Skewness of response latency distributions: A positive skewness is everywhere indicated, but it is predicted that the amount of skewness decreases as learning progresses.

It is concluded that these predictions are supported qualitatively by data that have already been collected. A large number of investigators have found that a response becomes quicker and more certain as learning progresses and have found asymptotes for both the central tendency and variability of latency distributions. Similarly, many investigators have found positive skewness in latency distributions. Thus, the gross features of the present theory are in agreement with past observations.

The desirability of a more precise test is pointed out and the implications for such a test of certain assumptions necessary to the theoretical development are discussed at some length. The required controls are shown to be mutually contradictory so that some compromise is necessary in the experiment. A statistical test for the selection of data in which the compromise has been adequate is briefly described. It is indicated that this test is independent of the theory which is to be tested.

An experimental test of the theory developed in this paper has been made and will be reported at an early date.

MATHEMATICAL NOTES

(1) *Notation and basic functions.* To avoid the writing of complicated expressions, functional abbreviations will be used. These abbreviations are set forth as they occur in particular equations. In general, the symbols g and G are reserved for distribution functions (the oscillation and variability functions); the small letter denotes the non-cumulative form and the capital letter the cumulative. The symbol f denotes any other function occurring in the theory. Within this scheme, the particular functions will be discriminated by means of subscripts. The symbol ϕ is used for the probability integral.

The equation for the growth of effective reaction potential with successive

reinforcements is:

$$\bar{E} = E_0 + (M_E - E_0)(1 - e^{-iN}) \equiv f_1(N). \quad (1)$$

The relation between t and $\hat{E}$ may be written as:

$$t = t_0 + \frac{a}{\hat{E} - L} \equiv f_2(\hat{E}). \quad (2)$$

If (2) is solved for $\hat{E}$, there results:

$$\hat{E} = L + \frac{a}{t - t_0} \equiv f_3(t). \quad (3)$$

The non-cumulative oscillation function of $\hat{E}$ is given by:

$$g_1(\hat{E}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\hat{E} - \bar{E} + 2.5\sigma)^2}. \quad (4)$$

The cumulative oscillation function is given by:

$$G_1(\hat{E}) = \int_{-\infty}^{\hat{E}} g_1(\hat{E})d\hat{E}, \quad (5)$$

which becomes, upon substitution of (4) and integration,

$$G_1(\hat{E}) = \phi\left(\frac{\hat{E} - \bar{E} + 2.5\sigma}{\sigma}\right). \quad (6)$$

(2) *The variability function for t.* The non-cumulative variability function is derived from (3) and (4) as (**6**, p. 24):

$$g_2(t) = g_1[f_3(t)]\left|\frac{df_3(t)}{dt}\right|. \quad (7)$$

By making the indicated substitutions from (3) and (4), the non-cumulative variability function in terms of t is:

$$g_2(t) = \frac{a}{(t - t_0)^2}\frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{1}{2\sigma^2}\left(\frac{a}{t-t_0} + L - \bar{E} + 2.5\sigma\right)^2}. \quad (8)$$

It is important for a later discussion to note that $g_2(t)$ can also be written as:

$$g_2(t) = \frac{a}{(t - t_0)^2} g_1[f_3(t)]. \quad (9)$$

As in the case of the oscillation function, the cumulative form of the variability function is obtained by integrating the non-cumulative form. Therefore, the cumulative variability function is given by:

$$G_2(t) = \int_{t_0}^{t} g_2(t)dt. \quad (10)$$

If (8) is substituted into (10) and the integration is carried out, there results:

$$G_2(t) = 1 - \phi\left[\frac{a}{\sigma(t - t_0)} - \frac{\bar{E} - L}{\sigma} + 2.5\right]. \quad (11)$$

Equation (8) is graphed on Figs. 3 and 5 and equation (11) on Fig. 4. An examination of the two equations shows that, although they are basically functions of t, they contain $\bar{E}$ as a parameter. Hence the shape of the curves must change as learning progresses (Compare Fig. 3 and Fig. 5).

At the asymptote, where $\bar{E} = M_E$ (11) becomes:

$$G_2(t) = 1 - \phi\left[\frac{a}{\sigma(t - t_0)} - \frac{M_E - L}{\sigma} + 2.5\right]. \quad (12)$$

Equation (12) can be fitted to data from the asymptotic phase. It contains t_0, a/σ, and $(M_E - L)/\sigma$ as parameters.

(3) *The median value of t* (t_1). Since the median value is that value below which 50% of the measures lie, t_1 is defined by the expression:

$$G_2(t_1) = \tfrac{1}{2}. \quad (13)$$

Substitution of (11) into (13) yields, after rearrangement:

$$\phi\left[\frac{a}{\sigma(t_1 - t_0)} - \frac{\bar{E} - L}{\sigma} + 2.5\right] = \tfrac{1}{2}. \quad (14)$$

Equation (14) may be written as:

$$\frac{a}{\sigma(t_1 - t_0)} - \frac{\bar{E} - L}{\sigma} + 2.5 = \phi^{-1}(\tfrac{1}{2}), \quad (15)$$

where the symbol ϕ^{-1} denotes the inverse of the probability integral. Since $\phi(0) = \frac{1}{2}$, the inverse relationship $\phi^{-1}(\frac{1}{2}) = 0$, must hold. With this substitution, (15) becomes, when solved

for t_1:

$$t_1 = t_0 + \frac{a}{\bar{E} - L - 2.5\sigma}. \quad (16)$$

The substitution of (1) into (16) gives:

$$t_1 = t_0 + \frac{a}{f_1(N) - L - 2.5\sigma} \quad (17)$$

Equation (17) yields, when $f_1(N)$ is replaced by the expanded expression in (1), the relation between the median latency and the number of reinforcements. It is plotted on Fig. 6 for arbitrarily chosen values of the parameters.

(4) *The modal value of t* (t_2). Since the mode is defined as the value corresponding to the maximum of the distribution curve, t_2 is defined by the expression:

$$\left.\frac{dg_2(t)}{dt}\right|_{t=t_2} = 0. \quad (18)$$

If (8) is differentiated and the result set equal to zero, as indicated in (18), a quadratic equation in t which has both a positive and a negative root is the result. Since t is by definition restricted to positive values, the positive root of the equation is taken as t_2.[10] The resulting equation is:

$$t_2 = t_0 + \frac{\sqrt{[\bar{E} - L - 2.5\sigma]^2 + 8a^2} - [\bar{E} - L - 2.5\sigma]}{4\sigma}. \quad (19)$$

If substitution for $\bar{E}$ is made from (1), an expression relating t_2 and N is obtained:

$$t_2 = t_0 + \frac{\sqrt{[f_1(N) - L - 2.5\sigma]^2 + 8a^2} - [f_1(N) - L - 2.5\sigma]}{4\sigma}. \quad (20)$$

Equation (20) is plotted for arbitrary values of the parameters in Fig. 6.

(5) *The moments of the variability function.* It is convenient to discuss the moments of $t - t_0$ rather than the moments of t. Any moment of $t - t_0$ differs from the corresponding moment of t by a finite quantity.

The rth moment of $t - t_0$ is given by (6, p. 36):

$$\mu_r = \int_{t_0}^{\infty} (t - t_0)^r g_2(t) dt. \quad (21)$$

If (9) is substituted into (21), there results:

$$\mu_r = \int_{t_0}^{\infty} a(t - t_0)^{r-2} g_1[f_3(t)] dt. \quad (22)$$

It can easily be shown that:

$$\lim_{(t-t_0)\to\infty} g_1[f_3(t)] \neq 0. \quad (23)$$

It is proved in function theory that an improper integral which can be written in the form (22) where the function occurring inside the integral has the property (23) will diverge if (2, p. 278):

$$2 - r \leq 1. \quad (24)$$

(24) is equivalent to:

$$r \geq 1. \quad (25)$$

For the mean the equality holds; for the variance and higher moments the inequality holds. Thus, the integral in (21) must diverge, and the moments of this distribution do not exist.

[10] The second derivative of $g_2(t)$ is negative at this point.

REFERENCES

1. FELSINGER, J. M., GLADSTONE, A. I., YAMAGUCHI, H. G., & HULL, C. L. Reaction latency (st_R) as a function of the number of reinforcements (N). *J. exp. Psychol.*, 1947, **37**, 214–228.
2. FRANKLIN, P. *A treatise on advanced calculus.* New York: John Wiley and Sons, 1940.
3. GLADSTONE, A. I., YAMAGUCHI, H. G., HULL, C. L., & FELSINGER, J. M. Some functional relationships of reaction potential (sE_R) and related phenomena. *J. exp. Psychol.*, 1947, **37**, 510–526.
4. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. ——, FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. PSYCHOL. REV., 1947, **54**, 237–254.
6. WILKS, S. S. *Mathematical statistics.* Princeton: Princeton University Press, 1944.

[MS. received February 11, 1949]

COLOR VISION AND FACTOR ANALYSIS

BY JOZEF COHEN

University of Illinois

It is not widely recognized that modern color theory and multiple factor analysis have much in common. This situation is not surprising since workers in one field are usually not interested in or conversant with the other. Though the differences between the two fields are many, the similarities are striking.

Color theory proposes to interpret an infinite number of color sensations in terms of a finite number of primary colors. Factor analysis proposes to interpret an infinite number of test-scores in terms of a finite number of factors or primary mental abilities. In both cases such an attempt is wise, since it is the duty of science to predict an infinite number of observations in terms of a finite number of variables; this aids in understanding the phenomena, since economy of thought is acquired in the process.

The fundamental equation of factor analysis is written

$$s_{ji} = c_{j1}x_{1i} + c_{j2}x_{2i} + c_{j3}x_{3i} + \ldots + c_{jq}x_{qi},$$

which means that the standard score of individual i on test j can be expressed as a linear function of q variables, in this case factors. The x's indicate the standard scores of the same individual j on each of the factors, and the c's are weights assigned to each of the factors. The number of factors, q, may be any number equal to or greater than unity, but it is hoped that it will be very small. The fundamental equation of color theory is written

$$C = aR + bB + cG,$$

which means that any color seen, C, is a linear function of three variables, here primary colors designated, for convenience only, as R, G, and B. The metric of a, b, c will be discussed below. This equation is an experimental fact.

The test space and the corresponding S matrix are generated by testing N individuals with n tests. This gives an $N \times n$ matrix which, after suitable normalizing of the cells, when multiplied by its transpose, gives a matrix of scalar products, termed a correlation matrix. This correlation matrix is the *only* invariant under rotation of the axes in the orthogonal group (**52**, p. 32). If there is any sense to the factorial problem, the test vectors (and, of course, the test scores) will lie in a space whose dimensionality is less than the order of the correlation matrix. This subspace is termed the common factor space, and contains no natural position for the new reference axes.

The color space is generated by no arbitrary selection of tests, but rather by a selection of real colors which are physically available. The generation of the space or the gauging of the spectrum (*die Eichung des Spektrums*) is accomplished by selecting any three colors (with only minor exceptions they may be spectral or non-spectral, saturated or unsaturated) and matching them against a population of spectral colors selected so as to cover the entire visible range, say colors that differ from each other by 10 millimicrons. The population must be of spectral colors, since less saturated colors would not determine the outer bounds of the space. Any three colors not on a

straight line on the chromaticity diagram, to be defined below, will serve as primaries, but all colors cannot be matched by any suitable selection of real primaries.[1] All colors can be specified in terms of three primaries, however, by allowing negative coefficients in the fundamental color equation. That is, in certain cases, the observer will be unable to match a given color by a mixture of the three primaries. If we allow him to mix one of the primaries with the color to be matched, then he will be able to match that mixture with the remaining two primaries. The first primary now assumes a negative coefficient. The order of the correlation matrix may be any number, the rank of the correlation matrix may be any number equal to or less than the order, but the order and rank of the spectral matrix is *always* three. A unique match to any color cannot be made by using more than three primaries, and any concept of a tetrachromatic gauging of the spectrum, as was suggested, for example, by Murray (36) is ill advised. The data from a tetrachromatic gauging would be infinitely long, and if desired for some reason or other, could be predicted from the trichromatic gauging data.

There is no such thing as a correlation coefficient between two colors or between two primary colors. The correlation coefficient in this case is not invariant under rotation of axes because the color vectors do not constitute an orthogonal group. This will be clear if one realizes that the correlations between the spectral primaries will change depending upon the original selection of the primaries. Consider a spectrum gauged with the primaries A, B, and C. Consider a population of individuals who take tests A′, B′, and C′. We could conceivably correlate A and B. We could also correlate A′ and B′. If we substitute primary D for C, and regauge the spectrum, the correlation between A and B would in general change. The correlation between A′ and B′ would never change no matter what we substituted for C′. The correlation matrix over the color space is absolutely meaningless.

The reduction in rank in the factorial problem is accomplished by factoring the correlation matrix and using some criterion for determining when sufficient factors have been extracted. This gives an F matrix, $n \times q$, where n is the number of tests and q is the number of factors. The spectral matrix, on the other hand, is derived directly from the observations and is $3 \times$ the number of colors selected from the spectrum. In the factorial problem, any F matrix will reproduce the correlation matrix but not the original S matrix. In the color problem, any spectral matrix allows unique specification of any color in the color space.

The difficulty in the analysis of both problems lies here. In the factorial problem, it is the location of the axes in the sub-space, and in the color problem, it is the location of the axes in the original color space, that factor theories differ among themselves and that color theories differ among themselves. In each case there is no intrinsic criterion for the position of the axes. Several philosophical arguments are presented in both instances. Should the factors and the primaries correspond to the "real" factors of the mind and to the "real" receptors of the eye? Is it possible to have a legitimate bi-polared

[1] The statement that all colors can be produced by the appropriate selection of primaries is a myth that ought to be cleared up among our textbook writers. For example, in C. T. Morgan's *Physiological Psychology* (34, p. 207) we find the statement, "Any kind of color experience can be produced if we select three appropriate wave lengths and mix them in different proportions."

factor, or a primary receptor which responds both positively and negatively?

The factor analysts use many criteria. Thurstone (**49**, pp. 319–346) selects his axes so that the *F* matrix will have a maximum of zero loadings and a minimum of negative loadings. This is an oblique structure—termed a simple structure. Hotelling (**15**) prefers a configuration called principle axes satisfying the requirements that the sums of the squares of the projections shall be at a maximum or at a minimum, as the case may be. The structure is orthogonal. Holzinger and his students (**14**, pp. 111–154), on the bases of the characteristics of the tests which are intercorrelated, want a "hollow stair case" pattern, which has characteristic general factors and group factors. Cattell (**5**) has recently introduced a new criterion of "parallel proportional profiles"—an extension of Thurstone's simple structure, wherein factorial invariance from test battery to test battery is sought. Other solutions have been offered by Spearman (**46**), Burt (**3**), Kelley (**24**), and Tryon (**51**).

The spectral matrix has certain characteristics which aid in the selection of the primaries. When plotted by rows, the color space is generated. But since each row corresponds to a given wave length, and the wave lengths lie on a continuum, then plotting the matrix by columns gives a most interesting configuration. If the points are connected, we derive a plot of the so-called "fundamental sensation curves," examples of which are found in most elementary psychology textbooks. It was at one time thought that the position of the axes could be determined by the appropriate selection of the fundamental sensation curves, so that when multiplied by the appropriate constants, termed *luminosity coefficients*, the weighted fundamental sensation curves would summate to the specified values of the differential sensitivity to wavelength curve of the eye. It was shown by Judd (**19**), however, that constants can be found for any set of curves so that the appropriate summation can be performed. But constants which have elegant characteristics and the nature of the curves themselves are important, as will be seen.

There are also many psychophysiological effects, not explained by the gauged spectrum, which aid in the selection of the primaries. Among these effects are color adaptation, after images, the flight of colors, simultaneous contrast, color blindness, the differential sensitivity of the visual fields, color limens, and the Prevost-Fechner-Benham subjective colors. The Helmholtz theory (**12**, pp. 141–146) is based on the fact that there are only brightness differences from 670 millimicrons on out to the long wave end of the spectrum. This selection of the fundamental sensation curves serves as the best possible explanation. The curves are not in disagreement with the known facts of color blindness. Recently, Szekeres (**48**) readjusted these primaries so that the fundamental sensation curves would be more symmetrical. Ladd-Franklin (**27**) based her theory on the measurements of König and Dieterici (**25**). Here the fundamental sensation curves have a foundation in the phylogeny of man—results of which exist now in the form of differential sensitivity zones in the retina. The Ladd-Franklin theory states that early in man's history, there were only white receptors. Later the white split to a yellow and a blue. Finally, the yellow split to a green and a red. There are now only three fundamental sensation curves (for an unemotional statement, see **27**, **21**).[2]

[2] The statement that Ladd-Franklin hypothesized a four-color theory is another myth perpetuated by our textbook writers. For example, Morgan (**34**, p. 207) writes, "The four-

Hering (13) utilized the phenomenon of after images in the selection of his fundamental sensation curves. There are three curves, two of which are bipolared. Hecht (11) selected his curves so that they would approximate the sensitivity curve of the eye, but not necessarily coincide with it. Wright (53) based his curves on the color adaptation characteristics of the eye. Stiles (47) has recently selected primaries based on the just noticeable increments of color stimuli of MacAdam (28, 29, 30), as did Moon and Spencer (32) before him.

The ICI system (45; for a report in an American journal, see 20), which was adopted by the International Commission on Illumination in 1931, was selected so that the luminosity coefficients, after the suggestion of Judd, be a simple 0, 1, 0, respectively. This means that one of the fundamental sensation curves (in this case a sensitivity curve, since the system is designed primarily for ease in physical specification) *is* the visibility curve, as is also the case, as a matter of fact, in the Hering theory. Among other rotations which have been suggested are those by Adams (1), Müller (35), Schrödinger (42), von Kries (26), and Moon and Spencer (33). All of these theories and their spaces are derivable one from the other by multiplying the spectral matrix by the appropriate transformation matrix—just as one solution of factor analysis is derivable from another by the appropriate rotation. For an excellent discussion of these transformations, see Judd (21).

Since the vectors of the F matrix are normalized, the sums of the squares of the factor loadings, including the unique factor, should always equal unity (for orthogonal frames). There is no similar function in the color space since such a relationship depends upon the orthogonality of the group. The sum of the coefficients in the color equation indicates the intensity of the color, while the ratio indicates the hue and saturation. The sum indicates the intensity in a metric called *trichromatic units*. The coefficients themselves are called *tristimulus values*. Obviously, the intensity specification of colors can be equated by allowing the sums of the tristimulus values to equal unity. When this is done, the new colors so derived are specified by so-called *trichromatic coefficients* which, by definition, always add to unity. Trichromatic coefficients specify colors which lie in a plane $x + y + z = 1$. When the colors on this plane are projected to the xy plane, a configuration is obtained which is termed the *chromaticity diagram*. There is a chromaticity diagram for every color theory of the type described above. It follows that there are five attributes of all these color theories: the spectral matrix, containing tristimulus values, which is the most powerful specification and from which all other attributes are derived; the color space which is the spectral matrix plotted by rows; the matrix of trichromatic coefficients, which is the normalized spectral matrix; the fundamental sensation curves which are the (normalized) spectral matrix plotted by columns; and the chromaticity diagram, which is the projection of the normalized tristimulus plane onto the xy plane.

All theorems of projective geometry for closed spaces apply to the projec-

color theory of Ladd-Franklin postulates, in addition to the three primary colors of Young and Helmholtz, a fourth type, the 'yellow' cone." One wonders why some writers insist that Ladd-Franklin postulates four cones as being present in man, when, as a matter of fact, if they were to carry this reasoning to its ultimate conclusion, they would expect to find *five* cones; white, the most primitive cone, would necessarily be included. Ladd-Franklin clearly states that yellow is a secondary process.

tive plane. All chromaticity diagrams are projections of the same space, but are distorted by the "pulling and pushing" of the rotation. The spectral colors and the purple line always lie at the periphery. The color white is theoretically at the point (.33, .33). All other colors are in the space, their physical saturation and hue being determined by their closeness to white and the hue line on which they happen to lie. Two colors, when mixed additively, will produce colors which lie on the line connecting them; the proportions will be in the inverse ratio of the segments into which the line is divided.

The unique characteristics of the color space and the accompanying normalized projections allow a somewhat specialized method of rotation, although basically it is no different than that of factor analysis. The transformations are reviewed by Cohen and Walker (**7**). It will be noted that there is no simple transformation to the psychological primaries of hue, saturation, and intensity. The psychological specifications of the Munsell and Ostwald systems order themselves in contours in the trichromatic color space. The contour lines for the Munsell system have been determined by Newhall (**37**), and Newhall, Nickerson and Judd (**38**), and for the Ostwald system by Foss, Nickerson, and Granville (**8**).

Methods of specifying a color of complex spectral characteristics as a point in the colorimetric space are discussed by Hardy (**10**, pp. 32–60), and recently by Robinson (**41**).

The metric of the factorial space is well defined when tests are inter-correlated, but it is ill defined when individuals are correlated against each other, as in inverted factor analysis. The metric of the color space proves to be a most difficult problem and resolves to heterochromatic photometry. What do we mean when we say that a given red is equal in intensity to a given green? It is interesting that the question should come up at all, and that we are able to render the judgment under any conditions, since the two sensations are entirely different. Adjustment is usually made so that equal proportions of the primaries will mix to give white. The additional important aspects of the metric problem are discussed by Wundheiler (**54**).

The sampling problem in factor analysis is not well understood, but it is best analyzed by Holzinger and Harman (**14**). Analogous to this problem is the difference limen in color perception. An excellent paper on this latter subject is by Nickerson and Stultz (**39**). Elaborate experiments by MacAdam (**29**, **30**) show that the limens are small ellipsoids in the 3-space and ellipses in the projected space. They are always ellipses since, as MacAdam (**28**) points out, ellipses and only ellipses always project as ellipses. These ellipses vary in size depending on their position in the color space, but their magnitude can be predicted.

Thurstone (**49**, pp. xiii–xiv) has suggested that more meaningful results to the factorial problem might be obtained by the use of non-metric approaches. There has been some work here, but it is as yet trivial. That the color space is non-Euclidean has been suggested by Troland (**50**, p. 250), and explicitly stated by Silberstein (**43**, **44**). There have been no successful non-Euclidean approaches to the color problem.[3]

There have been several references

[3] In a recent paper by A. E. Jacobson (**17**) it appeared that it had been demonstrated experimentally that the ICI color space was not ideal. In Judd's paper (**23**), which followed, it was shown that compensation for the deficiency in the ICI system could be made by adjustment of the data on the standard observer, and the Euclidean properties of the space remained intact.

to the color problem in factor analytical literature. The first is by Thurstone, who writes (49, p. 54):

". . . when the relations are preferably non-linear and mathematically involved, it is frequently possible to use the simple linear forms as first approximations. A well known example of this type of relation is that in which the chroma of a spectral color is expressed as a linear function of two arbitrarily chosen primaries. If two spectral colors are chosen arbitrarily for use as primaries, it is possible to express any immediate color as a linear function of the two arbitrarily chosen primaries. The coefficients of the two terms of this linear function represent the angular sizes of the two sectors into which a color rotator is divided. When the rotator is spun, the intermediate color is seen. But here . . . although the chroma of the resulting color is expressed in terms of the linear function of the arbitrary primaries, it does not follow that the saturation and the gray values are expressed by the same law. There is still debate about which colors are to be considered primary. The question can be settled only by discovering that a certain set of primaries give the most parsimonious comprehension of some phase of color vision." [4]

It should be pointed out that these relationships stated by Thurstone hold only when certain restrictions are placed on the primaries. If they are chosen so that the line connecting them on the chromaticity diagram passes through a region of the spectrum where the change in hue is approximately linear, then the first approximation would be very good. On the other hand, if primaries are selected so that the line connecting them on the chromaticity diagram passes through a region of the spectrum where the change in hue is non-linear, as in the green region of the ICI diagram, then the approximation would be inaccurate. A similar analysis is true for intensity and saturation.

The second reference to color in the factor analytical literature is by Allport and Odbert (2, p. 34). Herein are listed no less than 17,953 trait names. Allport and Odbert indicated that since this large number of words could not describe personality completely, what was needed was some additional names. In their efforts to find analogous situations in science, they invaded the field of color with the statement, "There are perhaps 300 words in English designating color experience.[5] By tetrachoric correlations, this list might be greatly reduced, but what an inadequate vocabulary would result for the profusion of distinguishable color sensations, estimated to exceed 2,000,000 in number!" To this Guilford (9) replied that factor analysis would give some ordering to personality just as it would to the color problem. The 2,000,000 names for colors would not only be highly confusing, but would tell nothing about how one color is related to another. Reference to three axes would give three numbers, uniquely specifying the color. Further, the numbers would tell the relationships between the colors so specified. Guilford suggests that if the analysis were performed, the primaries found would be a black-white, blue-yellow, and red-green factors. It is interesting to note that Guilford correctly anticipated the fact that hue, saturation, and intensity would not result.

Burt (4, pp. 20, 83–84), in 1940 called attention to the fact that color vision and factor analysis are closely related. Burt speaks of a factorial study on color vision by P. H. Chatterji being conducted in his laboratory, but

[4] Chroma in Munsell notation means saturation, but apparently Thurstone is using it to mean hue, as is sometimes done. By gray-value is meant value or intensity.

[5] *A Dictionary of Color* (31, pp. 189–207) lists approximately 3200 color names.

apparently it has never reached the literature.

The next reference to color is by R. W. Pickford (40), which describes briefly several experiments concerning the "factorial analysis of colour vision." Here Pickford *correlates* "tests" with each other, and *correlates* monochromatic and purple filters with one another. He tests various populations of normal and color deficient individuals, and factors the resulting matrices. The results are too many to reproduce here, except that he isolates three factors, a general, a red-green, and a yellow-blue. It is difficult to understand, in light of the arguments presented in the paragraphs above, how meaningful correlation matrices could have been generated by the techniques described by Pickford. This is not at all made clear in the short paper.[6]

The final reference to color and factor analysis is by Jones (18). It will perhaps be of interest to make a detailed commentary on this paper. In his opening paragraphs Jones proposes to introduce the techniques of factor analysis to the problems of psychophysiology in the hope that they will either give new results or confirm old ones. He "factors" the visibility curve, obtains his results, and writes that since it was successful in this instance, "It is to be concluded that factor analysis deserves further application to psychophysiological problems" (18, p. 396). It is hard to see why a technique which aids in describing the eye *necessarily* aids in describing the ear, for example. They are entirely separate sense departments and interpret their stimuli most differently. The eye never resolves a stimulus into its component wavelengths, while the overtones in an auditory stimulus are easily heard. This seems to be an error of inductive reasoning, and Jones' conclusions can only refer to the applicability of factor analysis to color vision.

Essentially, Jones takes the visibility data of Coblentz and Emerson (6), "correlates" one wavelength with another and derives three factors which presumably are indicative of fundamental sensation curves. Now we *know* the relationship between the visibility curve and the fundamental sensation curves. The relationship is an experimental fact first introduced by Ives (16). His paper says, as we have indicated before, that the curves which embody the color mixture data when weighted by the appropriate luminosity coefficients will summate to the normal visibility curve. These luminosity coefficients will exist for any set of color mixture data, and are therefore not an additional restriction (see Judd, 19). Since these luminosity coefficients can be 0, 1, and 0 respectively, it is possible, as we have pointed out before, for the luminosity curve to be one of the fundamental sensation curves. Since this one curve is only one third of the whole number of fundamental sensation curves about which the factor analysis is supposed to be indicative, it would appear *a priori* that a factor analysis of the visibility curve could lead to no meaningful results. Nevertheless, the factor analysis[7] of

[6] A personal communication to Pickford several years ago requesting information brought the answer that the matrices or further details were not then available. It is hoped that eventually they will appear in the literature.

[7] There are several matters concerning the factor analysis in this paper which are quite separate from the theoretical implications set forth above, but which should be indicated. There is something the matter with the loadings of 559 millimicrons in Table IV. The sum of the squares of the loadings does not equal the communality. Since the communality is the same in Table II (before rotation), the error must be in the loadings. There is no specification of the position to

the correlations has been performed, and it will serve as an empirical test on the theoretical considerations of the present paper.

As we say, Jones has derived three factors which, when plotted by columns, are supposed to indicate the nature of the fundamental sensation curves. Now Jones has gone to the literature and found three curves by Judd (22) which supposedly are those best described by Jones' three factors. Jones gives some discussion to show that each of Judd's curves corresponds to one of the factors if one allows some slight adjustment of the factors up or down the baseline. There can be no doubt that Jones feels that his factors lead to Judd's sensation curves, since he has reproduced Judd's curves, has shown where each of his own corresponds to one of Judd's, and even uses the word "factor" to describe one of Judd's curves.[8]

Now the interesting aspect about the analysis is this—the luminosity coefficients of Judd's curves happen to be 0, 1, and 0 respectively, which indicates that one of his curves *is the visibility function.* Specifically, it is W_d, and even visual examination of the curve will show this to be so. This means that Jones has, by the technique of factor analysis, found three curves, one of which is the original curve from which the three were derived. If Jones were to factor curve W_d by his technique he would get curve K, curve W_p, and interestingly enough, curve W_d all over again. Apparently, this is what Jones has done. Of course, it is impossible to do this, since one cannot obtain what is physically not available. It would therefore seem that Jones' conclusion should have been that factor analytical techniques applied to a correlation matrix have no meaning in color theory.

This paper has shown the striking similarities of the theories of color vision and factor analysis; it has been further demonstrated that the correlation matrix over the color space is variant, and there is, therefore, no *direct* application of conventional factor analysis to the data of color mixture.

which the centroid matrix was rotated, although it probably was to as nearly a simple structure as could be obtained with an orthogonal frame.

[8] We do not understand why one would want to see a relationship between these two sets of curves in any event. After all, Judd's curves are describing the color-blind color space, while the Coblentz and Emerson visibility curve presumably defines a normal, standard observer.

REFERENCES

1. ADAMS, E. Q. A theory of color vision. PSYCHOL. REV., 1923, **30**, 56–76.
2. ALLPORT, G. W., & ODBERT, H. S. Trait-names. *Psychol. Monogr.*, 1936, **47**, No. 211, 34.
3. BURT, C. Factor analysis by sub-matrices. *J. Psychol.*, 1938, **6**, 339–375.
4. ——. *The factors of the mind.* New York: Macmillan, 1940.
5. CATTELL, R. B. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 1944, **9**, 267–283.
6. COBLENTZ, W. W., & EMERSON, W. B. Relative sensibility of the average eye to light of different colors and some practical applications to radiation problems. *Sci. Pap. U. S. Bur. Standards*, 1918, **14**, 167–236.
7. COHEN, J., & WALKER, R. J. The transformation matrix of color mixture data. *Amer. J. Psychol.*, 1946, **59**, 299–305.
8. FOSS, C. E., NICKERSON, D., & GRANVILLE, W. D. Analysis of the Ostwald color system. *J. opt. Soc. Amer.*, 1944, **34**, 361–381.
9. GUILFORD, J. P. Unitary traits of personality and factor theory. *Amer. J. Psychol.*, 1936, **48**, 673–680.
10. HARDY, A. C. *Handbook of colorimetry.* Cambridge: The Technology Press, 1936, pp. 32–60.
11. HECHT, S. Vision: II. The nature of the photoreceptor process. In *A handbook of general experimental psychology.* Worcester: Clark University Press (C. Murchison, Ed.), 1934, pp. 704–828.

12. VON HELMHOLTZ, H. *Treatise on physiological optics* (J. P. C. Southall, Ed.). Menasha: George Banta Publ. Co. for Optical Society of America, 1924 (3rd ed.), Vol. II, pp. 141–146.
13. HERING, E. Grunzüge einer Theorie des Farbensinnes. Sechste Mittheilung. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften.* Vienna, 1874, **69** (Abtheilung 3), 169–204.
14. HOLZINGER, K. J., & HARMAN, H. H. *Factor analysis.* Chicago: Univ. Chicago Press, 1941, pp. 111–154.
15. HOTELLING, H. Analysis of a complex of statistical variables into principle components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.
16. IVES, H. E. The transformation of color-mixture equations from one system to another. *J. Franklin Inst.*, 1915, **180**, 673–701.
17. JACOBSON, A. E. Non-adaptability of the I.C.I. system to some near-whites which show absorption in the blue region of the spectrum. *J. opt. Soc. Amer.*, 1948, **38**, 442–444.
18. JONES, F. N. A factor analysis of visibility data. *Amer. J. Psychol.*, 1948, **61**, 361–369.
19. JUDD, D. B. Reduction of data on mixture of color stimuli. *Bur. Stand. J. Res., Washington*, 1930, **4**, 515–548.
20. ——. The 1931 I.C.I. standard observer and coordinate system for colorimetry. *J. opt. Soc. Amer.*, 1933, **23**, 359–374.
21. ——. Color vision. In *Medical physics* (O. Glasser, Ed.), Chicago: The Year Book Publishers, 1944, pp. 265–282.
22. ——. Standard response functions for protanopic and deuteranopic vision. *J. opt. Soc. Amer.*, 1945, **35**, 199–221.
23. ——. A comparison of direct colorimetry of Titanium pigments with their indirect colorimetry based on spectrophotometry and a standard observer. *J. opt. Soc. Amer.*, 1948, **38**, 1093.
24. KELLEY, T. L. *Crossroads in the mind of man.* Stanford: Stanford Univ. Press, 1928.
25. KÖNIG, A., & DIETERICI, C. Die Grundempfindungen und ihre Intensitätsverteilung in Spektrum. *Z. Psychol.*, 1893, **4**, 241–347.
26. VON KRIES, J. Die Gesichtsempfindungen. In *Handbuch der Physiolgie des Menschen* (W. Nagel, Ed.). Braunschweig: Friedrick Vieweg und Sohn, 1905, Vol. 3, pp. 109–282.
27. LADD-FRANKLIN, C. The nature of the colour sensations. In *Treatise on physiological optics* by H. von Helmholtz (J. P. C. Southall, Ed.). Menasha: George Banta Publ. Co. for Optical Society of America, 1924 (3rd ed.), Vol. II, pp. 455–468.
28. ——. Projective transformations of color-mixture diagrams. *J. opt. Soc. Amer.*, 1942, **32**, 2–6.
29. MACADAM, D. L. Visual sensitivities to color differences in daylight. *J. opt. Soc. Amer.*, 1942, **32**, 247–274.
30. ——. Specification of small chromaticity differences. *J. opt. Soc. Amer.*, 1943, **33**, 18–26.
31. MAERZ, A., & PAUL, M. R. *A dictionary of color.* New York: McGraw-Hill, 1930, pp. 189–207.
32. MOON, P., & SPENCER, D. E. Metric for color space. *J. opt. Soc. Amer.*, 1943, **33**, 260–269.
33. ——. Metric based on the composite color stimulus. *J. opt. Soc. Amer.*, 1943, **33**, 270–277.
34. MORGAN, C. T. *Physiological psychology.* New York: McGraw-Hill, 1943, p. 207.
35. MÜLLER, G. E. Über die Farbenempfindungen. *Z. Psychol.*, 1930, Ergänzungsband 17–18.
36. MURRAY, E. Binocular fusion and the locus of 'yellow.' *Amer. J. Psychol.*, 1939, **52**, 117–121.
37. NEWHALL, S. M. Preliminary report of the O.S.A. subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1940, **30**, 617–645.
38. ——, NICKERSON, D., & JUDD, D. B. Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1943, **33**, 385–418.
39. NICKERSON, D., & STULTZ, K. F. Color tolerance specifications. *J. opt. Soc. Amer.*, 1944, **34**, 550–570.
40. PICKFORD, R. W. Factorial analysis of colour vision. *Nature, London*, 1946, **157**, 700.
41. ROBINSON, H. A. Evaluation of colorimetric tristimulus values by means of a new centroid method. *J. opt. Soc. Amer.*, 1946, **36**, 270–277.
42. SCHRÖDINGER, E. Ueber das Verhältnis der Vierfarben- zur Dreifarbentheorie. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften.* Vienna, 1925, **134** (Abtheilung 2a), 471–490.

43. SILBERSTEIN, L. Investigations on the intrinsic properties of the color domain. *J. opt. Soc. Amer.*, 1938, **28**, 63–85.
44. ——. Investigations on the intrinsic properties of the color domain. II. *J. opt. Soc. Amer.*, 1943, **33**, 1–10.
45. SMITH, T., & GUILD, J. The C.I.E. colorimetric standards and their use. *Trans. opt. Soc., London*, 1932, **33**, 73–134.
46. SPEARMAN, C. *The abilities of man.* New York: Macmillan, 1927.
47. STILES, W. S. A modified Helmholtz line-element in brightness-colour space. *Proc. Phys. Soc.*, London, 1946, **58**, 41–65.
48. SZEKERES, C. A new determination of the Young-Helmholtz primaries. *J. opt. Soc. Amer.*, 1948, **38**, 350–363.
49. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univ. of Chicago Press, 1947, pp. 319–346.
50. TROLAND, L. T. *The principles of psychophysiology.* New York: D. Van Nostrand, 1930, Vol. 1.
51. TRYON, R. C. *Cluster analysis.* Ann Arbor: Edwards Bros, 1939.
52. WEYL, H. *The classical groups.* Princeton: Princeton Univ. Press, 1939, p. 32.
53. WRIGHT, W. D. The measurement and analysis of colour adaptation phenomena. *Proc. roy. Soc.*, 1934, **115B**, 49–87.
54. WUNDHEILER, A. W. On the metric of the color space. *J. opt. Soc. Amer.*, 1946, **36**, 288–291.

[MS. received February 24, 1949]

HOMEOSTASIS AND SERVO-MECHANISMS FOR WHAT?

BY CHRISTIAN O. WEBER

Wells College

There is one respect in which organic principles may be extended to psychological processes without getting into the "mud of metaphysics" (as Rudolph Carnap says) which surrounds the body-mind problem. This extension consists in approaching behavior from the functional, or more boldly, from the teleological point of view. The organism and its mind, whether substantially different or not, alike dwell in an alien external world, contacts with which are prevailingly threats against or promises of survival. Hence adaptation and emergency action are the common principles of biological and psychological science. It is in this sense that Cannon's physiological principle of homeostasis and the more physical concept of servo-mechanisms have been successfully applied to aspects of behavior.

Granting the soundness of the extension of homeostasis and servo-theory as solvents for some psychological problems, it is the object of this paper to note one possible and general limitation of these concepts. As I shall indicate, Cannon himself was fully aware of this limitation in the case of homeostasis. After their full application to psychology the two principles in question would still leave much of human behavior unexplained. Perhaps criticism may even suggest that these principles do not extend beyond the physiological —they leave off where normal mental phenomena really begin. Both the concepts of homeostasis and of servo-theory are so new, at least in their attempted application to mental processes, that criticism of their use in psychology may seem premature. It may nevertheless be well to note in advance what they must undertake if they are to be successful general theories of behavior.

The limitation of these theories which the author has in mind is not specific and has no neat designation in current terminology. To state it briefly in advance, one may say that homeostasis and servo-theory do not answer the question, "Homeostasis and servo-mechanisms for what?" Consciousness appears to be the correlate of homeostasis upset and of servo-mechanisms which fail. These principles cannot account for the formulated ends of behavior of which we are so keenly aware, namely to strive and create as individuals. To attain these aims we perpetually upset homeostasis and continually create new servo-mechanisms. These criticisms, elaborated in what follows, will not apply where the concepts in question are not defended as *general* theories of all behavior. To make the difficulties inherent in the generalization of these views clear, this paper (1) begins with brief statements of the two related principles in question, (2) then briefly indicates and evaluates the ways in which homeostasis has been applied to mental processes, (3) presents Cannon's own versions of the limitations of homeostasis, and (4) gives some summary considerations.

Homeostasis and Servo-Mechanisms

Since W. B. Cannon's main published work on homeostasis is already ten years old (3), a brief indication of the meaning of this term will suffice. Homeostasis refers to the dynamic tend-

ency of physiological processes to maintain constant states in the open system of the organism. Cannon rejects the term *equilibria* as a proper designation for these states. The term *equilibrium* is traditionally applied to "relatively simple physico-chemical states, in closed systems, where known forces are balanced" (3, p. 24). Authoritative, clear, and engaging as always, Cannon shows the sweeping scope of homeostasis in "the wisdom of the body"—the maintenance of constant temperature, the chemical balances of the blood and numerous other bodily fluids and secretions. Homeostatic balancing is secured by two primary processes: (1) by storage, *e.g.*, as of salt, glucose, proteins, fat, calcium; and (2) by overflow of excess, as when excess glucose is released through the action of the kidneys. Apparently viewing these processes in the light of his famous "emergency" principle, Cannon dwells on the organism's "margins of safety." The heart, for instance, is able to double its rate, sending forth twice the usual volume of blood and increasing the arterial pressure by 30 or 40 per cent (3, p. 233). Having double organs (cerebral hemispheres, eyes, ears, kidneys, etc.) is another safety device making normal homeostasis possible. The organism's resistance to damage similarly points to the constant levels of functioning, as when renal functions are maintained despite the removal of two-thirds of each kidney (3, p. 234).

The concept of servo-mechanisms lies close to that of homeostasis in its central idea of a process which tends to return to normal tempo and direction after deviation. Servo-theory is interesting in that, unlike homeostasis, its mechanisms may exist in purely physical systems and yet can be extended to behavioral processes which are more purely psychological than the organic steady states discussed by Cannon. The regulation of heat, by means of the thermostat, or the control of the intake of steam into an engine by means of the centrifugal "governor" of Watt, are examples of purely physical servo-mechanisms. Such coördinating systems often depend on "feed-back" mechanisms, as when the manipulation of railway semaphore and rail switches results in return signals indicating that the intended adjustment has taken place. Norbert Wiener (11) and his associates have exploited the significance of this concept for understanding behavioral processes, such as the role of proprioceptive impulses in governing motor coördination. Kreezer (5) is exploring the use of servo-theory in understanding the normal psychology of drive, and Cameron (2) accepts the fruitfulness of this concept in mental pathology without qualification. Since, perhaps because of its greater age, the concept of homeostasis has been more frequently applied to psychological processes, the remainder of this paper will be confined to it.

HOMEOSTASIS AND MENTAL PROCESSES

J. M. Fletcher (4) indicates interesting and plausible ways in which homeostasis might account for a number of normal psychological processes: color vision, after-images of seen movement, resistance to mental distraction and to unfavorable temperature during work, perceptual constancy phenomena, and aspiration level. He also notes its special applicability in the sphere of personality maladjustment. D. W. Orr (8) introduces the concept of homeostasis into psychology through the avenue of instinct. The "instincts," broadly conceived as relatively autonomous organic drives appear to be native and unconscious mechanisms quite in keeping with Cannon's view of homeostatic processes. Orr concludes

that homeostasis may be regarded as instinctive and directed towards the attainment of security in a complex environment. It leads to conservative and orderly behavior and preference for the familiar. Regression is a way of attaining these stable states in the face of severe stress. In similar vein, L. S. Kubie (6) conceives instincts as demands made by the body on mental functions. R. M. Lindner (7), more concretely, finds analogies between organic homeostasis and the behavior of psychopaths. The patient guards against disintegration due to inner tensions by an overflow of energy, resulting in the familiar syndromes of hostility and destructive tendencies. Accepting this analysis, R. Van Vorst (10) notes the individual ways of restoring homeostatic balance, and attributes the consistently predatory and rebellious behavior of the psychopath to a weakening of restraining forces in his personality.

No doubt these extensions of homeostasis to mental processes have merit —as far as they go. To validate the psychological import of homeostasis through the unvalidated concept of instinct will be questioned by some. But impulses, whether called "instinctive" or not, often do appear in pairs, one nullifying the effects of the other (fear and anger, love and hate, etc.) as though homeostatic balance were their *raison d'etre*. Homeostasis may well be the common denominator of the protean ways in which "security" is sought in aberrant forms of behavior, and in this case has import for therapy.

However plausibly the symptoms of pathology may be interpreted as ways of seeking homeostatic stability, or as mechanisms which put higher mental processes in the service of organic drives, these views are particularly deflating when applied to normal individuals. Progress begins to look like an adjunct for preserving the archaic, and evolution is in the service of the *status quo*. The situation is nicely illustrated by Romer's (9) account of why the fins of certain fishes were modified so that they could walk on land. This, according to Romer, merely enabled them to get back into water again when the streams where they lived were dried up! Everything seems directed towards preserving life, and very little towards making it worth living. There is, of course, pragmatic common sense in this which daily behavior illustrates in endless ways. The mother who has her child taught how to swim reverses the aim of Romer's fishes, but the logic is the same—you do things in order to get back to where you were!

These views miss what, for want of a specific term, we might call the "instinct for adventure." The normal boy does not learn to swim through the shrewish motive of land-bound security. When an adult thinks, his sole motive is not to get rid of irritants which interrupted his favored state of mental coma. What of the scientist who never seems happy unless cognitive homeostasis is upset? P. W. Bridgman had the following to say about scientific method:

"The scientific method, as far as it is a method, is nothing more than doing one's damndest with one's mind, no holds barred" (1, p. 450).

Whatever the "aim" of the liver may be, a scientist equipped with his methods, which he perhaps loves more than anything he discovers by their means, is equipped to find life interesting. This interest illuminates and warms his life like a never-setting sun, while neural homeostasis continues in its chronic imbalance. To continue in this delectable state H. Poincaré was once driven to the extremity of using surveying instruments on roads which had already been surveyed. When chided for thus

using his time unprofitably, Poincaré replied that his problem was not to earn his meals but to keep from being bored between them. Scientists have no monopoly on this typically human characteristic. At this point the suspicion arises that homeostatic storage and balance are but advance provisions for future homeostatic loss for the sake of quite other ends. That this is the view of Cannon himself will appear in the section which follows. Those who import homeostatic theory into psychology will do well to study Cannon's views on this matter.

CANNON ON THE USES OF HOMEOSTASIS

Cannon's glimpse of vistas lying beyond homeostasis is perhaps embodied most concisely in the following passages:

"Repeatedly in foregoing chapters I have called attention to the fact that insofar as our internal environment is kept constant we are freed from the limitations imposed by both internal and external agencies or conditions that could be disturbing. The pertinent question has been asked by Barcroft, freedom from what? It is chiefly freedom for the activity of the higher levels of the nervous system and the muscles which they govern. By means of the cerebral cortex we have all our intelligent relations to the world about us" (3, p. 302).

"Bodily homeostasis, as we have learned, results in liberating those functions of the nervous system that adapt the organism to new situations, from the necessity of paying routine attention to the management of the details of bare existence. . . . With homeostatic devices, however, that keep essential bodily processes steady, we as individuals are free from such slavery—free to enter into agreeable relations with our fellows, free to enjoy beautiful things, to explore and understand the wonders of the world about us, to develop new ideas and interests, and to work and play, untrammeled by anxieties concerning our bodily affairs. The main service of social homeostasis would be to support bodily homeostasis. It would therefore help to release the highest activities of the nervous system for adventure and achievement. With essential needs assured, the priceless unessentials could be freely sought" (3, p. 323).

In broad outline, Cannon's view concerning the organic arrangements for securing these "priceless unessentials" is as follows: Homeostatic conditions are provided by many of the functions of the "involuntary" *interofective* (vegetative or autonomic) nervous system (3, pp. 249, 250, 261). It is the central nervous system, "voluntary" and *exterofective,* which utilizes homeostatic reserves, and produces organic fatigue (3, pp. 307–308). The dividend comes in the form of freedom for higher levels of the central system and the muscles which they control. Cannon notes that the motor controls of the cerebrospinal nervous system are almost wholly confined to muscles which control the bones (3, pp. 307–308).

In his "Epilogue" Cannon finds analogies to homeostatic storage in the social organism, as in the conservation of natural resources, and by numerous devices of taxation and insurance. These too provide "priceless unessentials" for the individual as for the group—new inventions and exploits, fresh ideas, even the "discords of human nature," and the "intrigues of love and hate." Why are these priceless although unessential? Because they make life varied and colorful, even as is personality itself, and provide for the "discipline and enjoyment of individual aptitudes" (3, p. 324).

SUMMARY CONSIDERATIONS

Cannon's implied distinction between organic necessities and "priceless unessentials" has its pitfalls. Many biologists, to be sure, credit the view

that there are activities of animals, especially human ones, which can persist despite their uselessness or even menace to homeostatic balance and survival. But the separation of necessary from "luxury" activities may be a consequence of sheer ignorance. Thus the poet Schiller, ignorant of the sexual activities of birds, declared that certainly their songs were free from the implications of want.

It is unnecessary to show that there are luxury activities which insure the enjoyment of life, existing apart from those which insure survival. Need and pleasure can be and usually are correlatives. Grant that the full gamut of man's capacities give him survival value and arose through natural selection. Grant the assumptions of Orr and Kubie that all mental activity is fundamentally in the service of organic demands. But why is it so seldom noted that the very processes which make survival possible include the very ones which *seen from the point of view of consciousness* make life significant and worth living? Eating, drinking, sleeping, and so on up to remembering, reasoning, creating, and appreciating: these activities may provide for survival, but for what reason would one wish to survive except to exercise these very functions? If life were not filled with these, with what would it be filled? Nature, with an unwonted but neat economy, has embodied means and ends in the same psychological activities. The situation seems to be that both homeostatic and conscious processes are the *causes* of survival, but only consciousness can be the source of those experiences which give meaning and worth to survival. As means for survival, organic homeostatic processes and overt behavior show harmonious alternation, like the proverbial hen and egg. Homeostatic storage requires appropriate overt behavior (food-seeking), which in its turn upsets homeostatic balance.

Homeostatic balance makes it possible to live at all but contributes little to living well. Homeostatic storage provides funds of energy, but does not specify the directions of its expenditure. In this respect, servo-mechanisms lie closer to behavior. Except perhaps for psychologists engaged in their paradoxical flight from the psychological, homeostatic balance is seldom a consciously formulated goal. Neurotics ridden with the obsession of insecurity may so formulate it, or the miser who makes "storage" an end in itself. The ridiculous character of homeostatic balance as a final goal is brought out in the story of the grocer who urged his customers to buy each of two kinds of breakfast cereal because one would stimulate them and the other one calm them down again. In short, having attained perfect homeostatic balance, and having nothing else to do, we would be driven to upset homeostasis itself, as Poincaré did, as a last desperate resort to keep in the stream of becoming. Perhaps the innumerable and useless activities which man is inclined to indulge must be explained in this way.

Consciousness seems to appear when homeostatic stability is in the process of attainment or loss, usually the latter. We need not dwell on the overwhelming evidence which shows that mental processes are "sparks" struck by *change*. This begins with the simplest sensations. Colors, odors, tastes die out as adaptation sets in, *i.e.*, as the homeostatic mechanisms upon which they rest reach stability. Even the normally unconscious processes of storage and overflow discussed by Cannon may invade consciousness when storage falls too far behind expenditure (fatigue), may give rise to feeling and impulse when they overflow (**7**, **10**), and appear as pain in various disorders. By

far the largest number of conscious and voluntary motor functions accompany an increasing homeostatic deficit. It is the cerebrospinal system which is always upsetting the apple cart of homeostasis. It controls man's legs, hands, and speech muscles, so that he can run into difficulties and then try to talk himself out of them again, meanwhile exhausting the sugar, fat, and salt balances which the tidy interofective system has built up. Since the latter system provides such generous reserves against calamity, we might suspect her collusion in the thirst for adventure which sums up so much of the central nervous system's activities.

Why is it that of all consciously experienced goal-seeking, the *ideals* of truth, beauty, and love are set above others? Part of the reason for this is that they are really not ultimate goals at all, but only *directions* in which behavior can move. They allow progress but can never be attained or perfected. They provide for an ever-shifting equilibrium, like the moments of time itself, attained only to be lost again. In short, ideals offer the best assurance of homeostatic balance perpetually falling short of attainment in the conscious sphere. The attainment of stable homeostasis is another name for dying. Cannon shows that the onset of old age is heralded by the increasing permanence of homeostasis attained. The physical inertia of a corpse is the point where homeostasis leaves off and the principle of equilibrium begins its reign.

BIBLIOGRAPHY

1. BRIDGMAN, P. W. The prospect for intelligence. *Yale Review*, 1945, **34**.
2. CAMERON, D. E. The current transition in the conception of science. *Science*, 1948, **107**, 553–558.
3. CANNON, W. B. *The wisdom of the body* (rev. ed.). New York: Norton, 1939.
4. FLETCHER, J. M. Homeostasis as an explanatory principle in psychology. PSYCHOL. REV., 1942, **49**, 80–87.
5. KREEZER, G. L. (Abstract.) *Amer. Psychologist*, 1948, **3**, 239.
6. KUBIE, L. S. Instincts and homeostasis. *Psychosom. Med.*, 1948, **10**, 15–30.
7. LINDNER, R. M. Psychopathic personality and the concept of homeostasis. *J. clin. Psychopathol.*, 1945, **6**, 517–521.
8. ORR, D. W. Is there a homeostatic instinct? *Psychoanal. Quart.*, 1942, **11**, 322–335.
9. ROMER, A. S. *Man and the vertebrates*. Chicago: Univ. of Chicago Press, 1941, **47**.
10. VAN VORST, R. Some responses of the psychopath as interpreted in the light of Lindner's suggested application of the concept of homeostasis. *J. clin. Psychopathol.*, 1947, **8**, 827–830.
11. WIENER, N. *Cybernetics, or control and communication in the animal and the machine*. New York: John Wiley, 1948.

[MS. received February 28, 1949]

REPORT TO LIBRARIANS

concerning

Available Back Issues

of the

Psychological Review

Thirty-two issues of the Psychological Review became out-of-print in 1948. Libraries and institutions ordered 11 sets of "all available issues," and 32 issues had fewer than 11 copies.

The American Psychological Association does not plan to reprint missing issues. At the rate of sale established for 1948, 24 more issues will become out-of-print during 1949.

The list of available issues of the Psychological Review, as found on the back cover of this journal, agrees with the inventory of December 31, 1948.

PSYCHOLOGICAL REVIEW

YEAR	VOLUME	AVAILABLE NUMBERS						PRICE PER NUMBER	PRICE PER VOLUME
1894	1	–	2	–	4	5	6	$1.00	$4.00
1895	2	–	–	3	4	5	6	$1.00	$4.00
1896	3	–	–	–	–	–	–	——	——
1897	4	1	–	–	–	–	6	$1.00	$2.00
1898	5	–	2	3	4	5	–	$1.00	$4.00
1899	6	–	–	–	–	–	6	$1.00	$1.00
1900	7	1	–	–	–	–	–	$1.00	$1.00
1901	8	1	2	–	–	–	–	$1.00	$2.00
1902	9	–	2	–	–	–	–	$1.00	$1.00
1903	10	1	2	3	–	–	–	$1.00	$3.00
1904	11	1	–	–	4	5	6	$1.00	$4.00
1905	12	1	2 & 3		4	5	–	$1.00	$5.00
1906	13	–	–	3	4	5	6	$1.00	$4.00
1907	14	1	2	3	–	5	–	$1.00	$4.00
1908	15	–	–	–	–	–	–	——	——
1909	16	1	–	3	4	5	6	$1.00	$5.00
1910	17	1	2	3	–	–	6	$1.00	$4.00
1911	18	1	2	3	4	5	6	$1.00	$5.50
1912	19	1	2	3	4	5	6	$1.00	$5.50
1913	20	1	2	3	4	5	6	$1.00	$5.50
1914	21	–	2	3	4	5	6	$1.00	$5.00
1915	22	1	2	3	4	5	6	$1.00	$5.50
1916	23	1	–	–	4	5	6	$1.00	$4.00
1917	24	–	–	–	4	–	–	$1.00	$1.00
1918	25	–	2	3	4	5	6	$1.00	$5.00
1919	26	1	2	3	4	5	6	$1.00	$5.50
1920	27	1	2	3	4	5	6	$1.00	$5.50
1921	28	–	2	3	4	–	6	$1.00	$4.00
1922	29	1	–	–	4	–	–	$1.00	$2.00
1923	30	1	2	3	4	5	6	$1.00	$5.50
1924	31	1	2	3	4	5	6	$1.00	$5.50
1925	32	–	2	3	–	5	–	$1.00	$3.00
1926	33	1	2	3	4	5	6	$1.00	$5.50
1927	34	1	2	3	4	5	6	$1.00	$5.50
1928	35	1	2	3	4	5	6	$1.00	$5.50
1929	36	1	2	3	4	5	6	$1.00	$5.50
1930	37	1	2	3	4	5	6	$1.00	$5.50
1931	38	1	2	3	4	5	6	$1.00	$5.50
1932	39	1	2	3	4	5	6	$1.00	$5.50
1933	40	1	2	3	4	5	6	$1.00	$5.50
1934	41	1	2	3	4	5	6	$1.00	$5.50
1935	42	1	2	3	4	5	6	$1.00	$5.50
1936	43	1	2	3	4	5	6	$1.00	$5.50
1937	44	1	2	3	4	5	6	$1.00	$5.50
1938	45	1	2	3	4	5	6	$1.00	$5.50
1939	46	1	2	3	4	5	6	$1.00	$5.50
1940	47	1	2	3	4	5	6	$1.00	$5.50
1941	48	1	2	3	4	5	6	$1.00	$5.50
1942	49	1	2	3	4	5	6	$1.00	$5.50
1943	50	1	2	3	4	5	6	$1.00	$5.50
1944	51	–	2	3	4	5	6	$1.00	$5.00
1945	52	1	2	3	4	5	6	$1.00	$5.50
1946	53	1	–	3	4	5	6	$1.00	$5.00
1947	54	1	2	3	4	5	6	$1.00	$5.50
1948	55	1	2	3	4	5	6	$1.00	$5.50
1949	56	By Subscription, $5.50						$1.00	——
								List price, Volumes 1 through 55	$241.50
								30% Discount	72.45
								Net price, Volumes 1 through 55	$169.05

Information about the Psychological Review: from 1894 to 1908, many numbers are out of print, as shown in the table. After 1908, twenty-three numbers are out of print. After 1925, two numbers are out of print.

The journal has been published with six numbers a year throughout its history.

Information about prices: the Psychological Review has the uniform price of $5.50 per volume and $1.00 per issue. For incomplete volumes, the price is $1.00 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Current subscriptions and orders for back numbers should be addressed to

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 Massachusetts Avenue, N. W. Washington 5, D. C.